National Aeronautics and
Space Administration

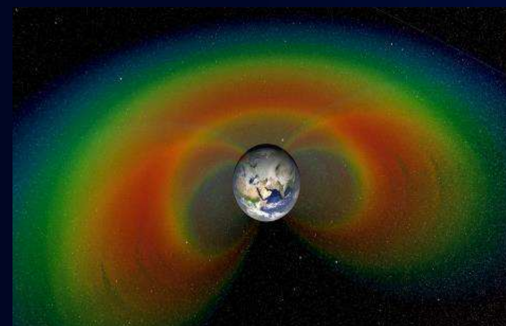
Advancing Space Radiation Biology with Big Data, Open Science, Standards, and ML

Sylvain V. Costes, Ph.D.
Space Biosciences Research Branch Chief
Project Manager for Open Science for Space Biology (GeneLab/ALSDA)
Lead Scientist for the Radiation Biophysics Laboratory
NASA Ames Research Center

Biologically Relevant Environmental Factors Encountered in Spaceflight

Microgravity/Reduced Gravity

Ionizing Radiation

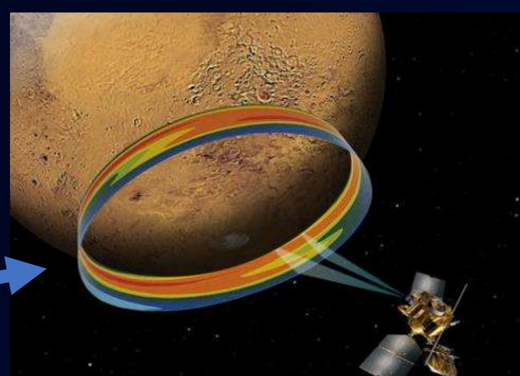


Credits: NASA/Goddard Space Flight Center/Scientific Visualization Studio

Altered Day/Night Cycles: Circadian Rhythm Changes

- Elevated CO2
- Reduced atmospheric pressure and elevated volumetric fraction of oxygen

Altered Temperature and Atmosphere



Isolation

COMBINATION OF MULTIPLE STRESSORS

Thriving in Deep Space (TIDES)

Biological effects of multiple **deep-space stressors**

1. Radiation
2. Gravity
3. Temperature & Atmosphere
4. Day & Night Light Cycles
5. Isolation
6. Regolith/dust

Cannot accurately replicate on the ground

Heavy ions can impact biological systems

Transformative biological science and exploration applications

1. Animal biology — Vertebrate and invertebrate models to probe analogous changes in humans
2. Plant Biology — From plant models to crops to sustain life for long-term human habitation
3. Microbiology — How it influences animals and plants in space

Thriving in Deep Space (TIDES)

Ground studies
Space studies
Ground & space studies

DEEP SPACE STRESSORS

Gravity | Radiation | Temperature/Atmosphere | Day/Night/Circadian Cycle | Isolation

PLATFORM PROGRESSION

Ground → Sub-Orbital → LEO/ISS → Gateway → Lunar Surface → Mars Transit → Martian Surface

Understand Fundamental Mechanisms

Use model organisms to determine fundamental biology of the human in the deep space environment

- Utilize "omics", systems biology, etc.
- Test small- & large-scale molecules, organisms, ecosystems
- Combined effects of radiation, gravity, & other stressors

Build the Blocks to Support Human Life

Study the environmental stressors in spaceflight on plants: crops and seeds

- Combined effects of radiation, gravity, & other stressors

Build a Foundation for Sustained Life on Mars

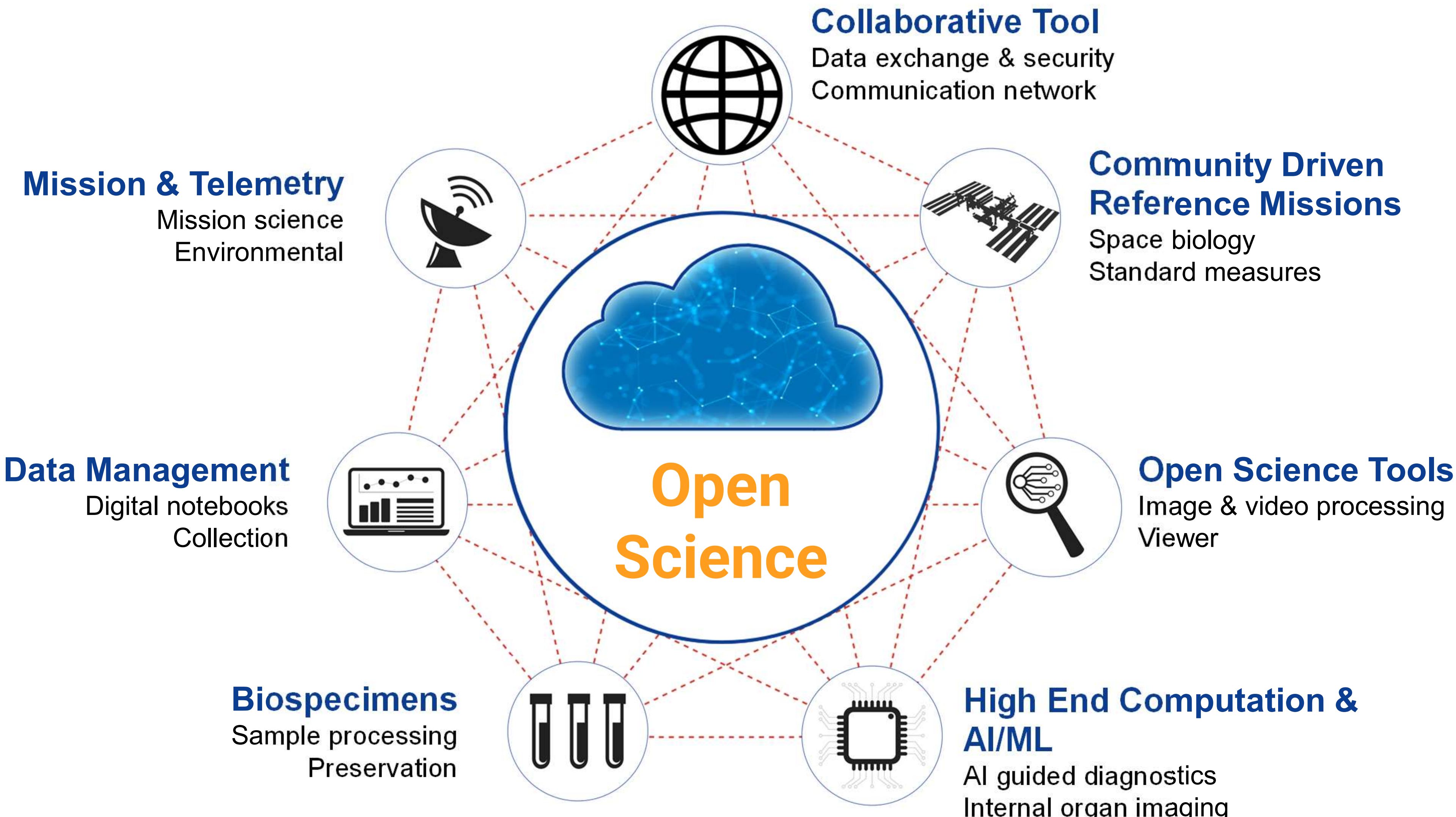
Engineer habitats and ecosystems to enable astronauts' independence from Earth

- Stabilize human/animal-plant-microbial ecosystems in the context of multiple deep space stressors

MODEL ORGANISM PROGRESSION

Unicellular (e.g., yeast, fungi) → Invertebrates (e.g., flies, worms, tardigrades) → Vertebrates (e.g., mice & rats; fish) → Humans

Unicellular (e.g., yeast, fungi) → Simple model plants (e.g., *Arabidopsis*) → Crops & edible plants → Humans


Collaborative Tool
Data exchange & security
Communication network
Mission & Telemetry
Mission science
Environmental
Community Driven
Reference Missions
Space biology
Standard measures
Open Science
Data Management
Digital notebooks
Collection
Open Science Tools
Image & video processing
Viewer
Biospecimens
Sample processing
Preservation
High End Computation &
AI/ML
AI guided diagnostics
Internal organ imaging

NASA Biological Open Science Resources

Biospecimen Sharing Program (BSP)

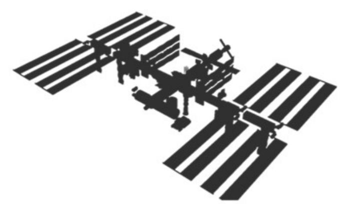
Dissection and preservation of rodent tissues from Flight and Ground investigations. Coordination of internal tissue sharing

NASA Biological Institutional Scientific Collection (NBISC)

Collection of non-human specimens and space microbial culture

Ames Life Sciences Data Archive (ALSDA)

Collection and curation of mission, project, and imaging data



NASA GeneLab (GL)

Collection and curation of omics data

NASA Internal Program

Open-Source Science Programs — Available Globally

433
Studies

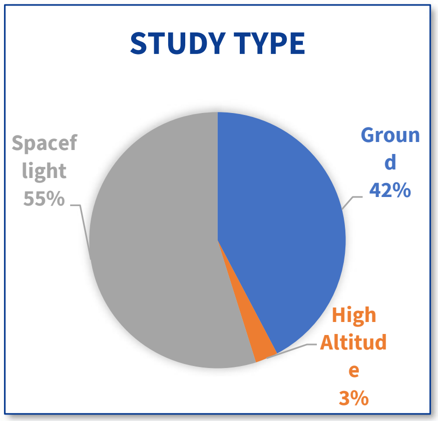
819
Datasets

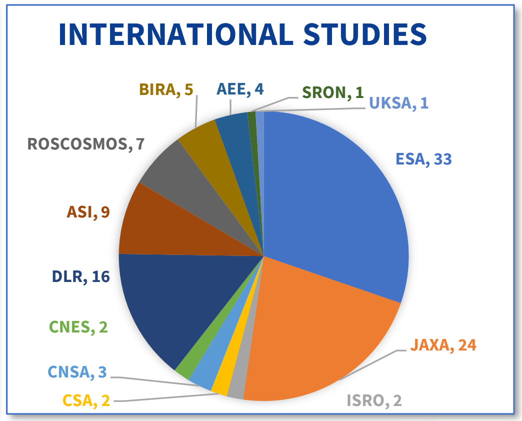
45
Species

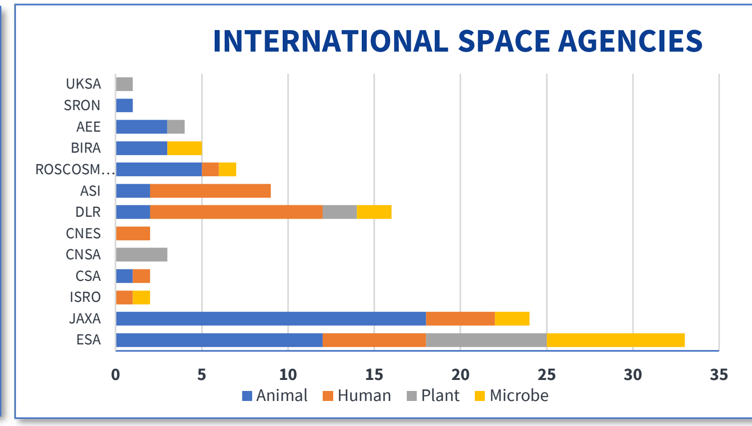
>25
Assays

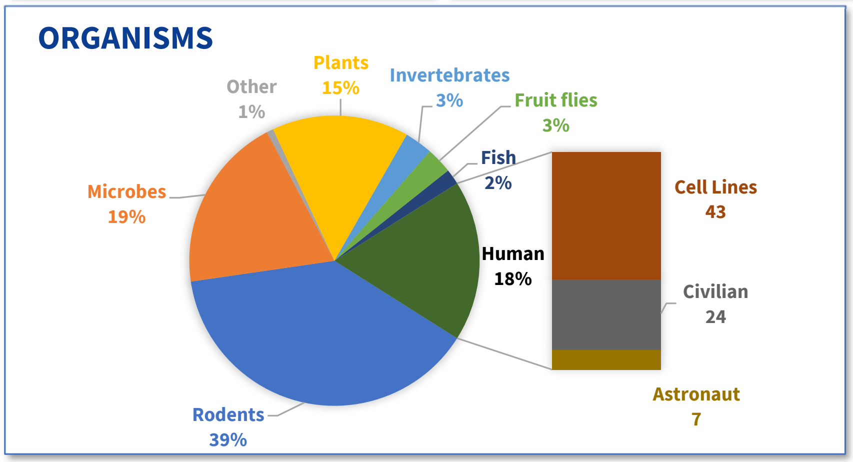
>150TB
Data

OSDR Database (GeneLab and ALSDA)

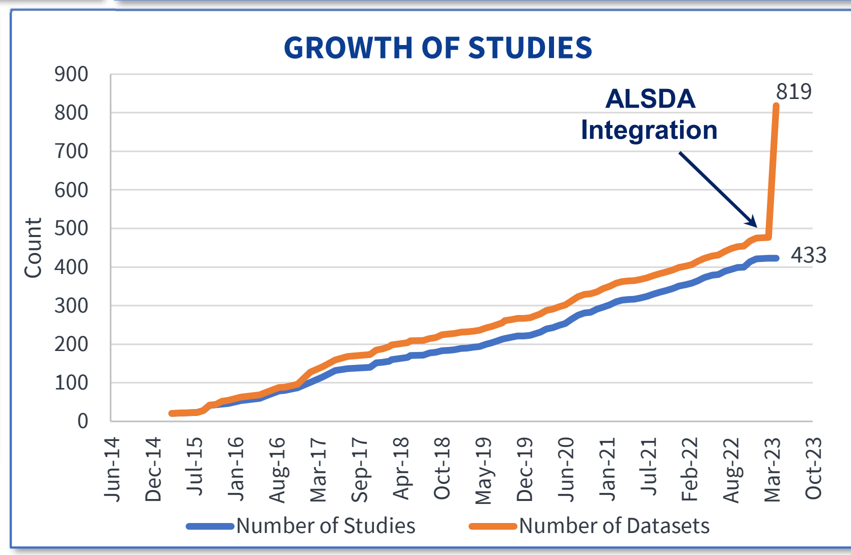








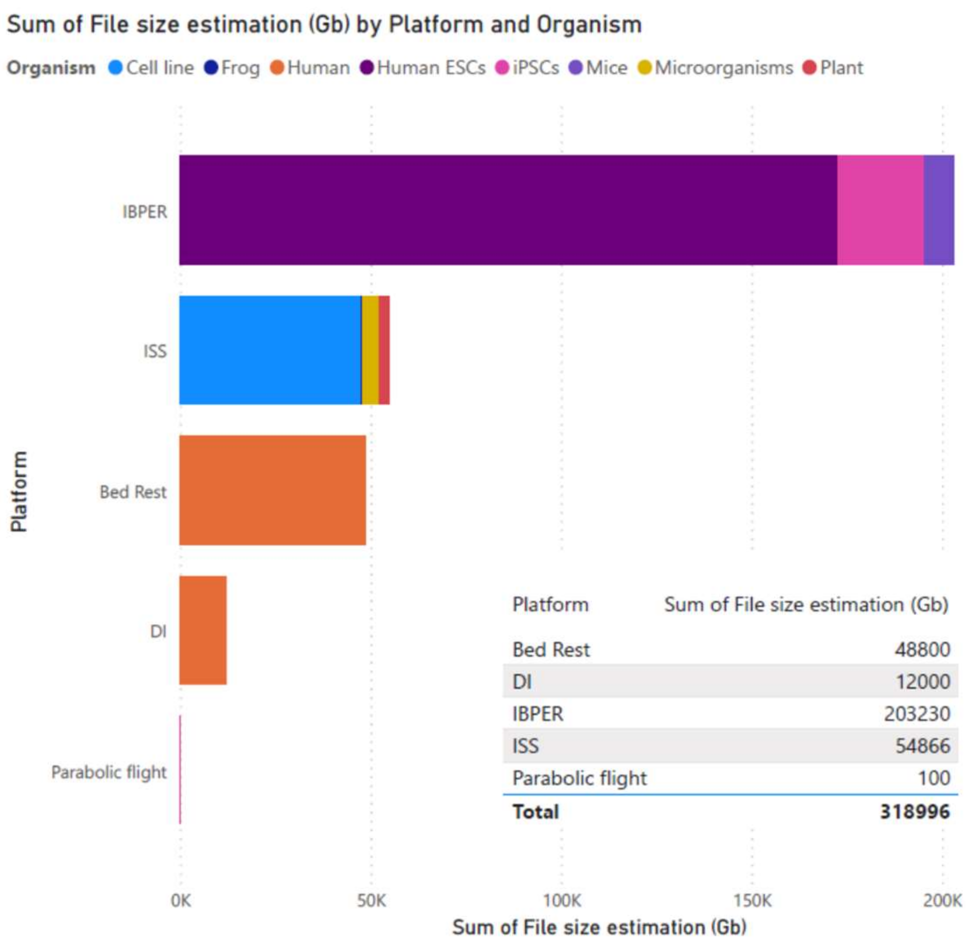
Civilian and Astronaut	Bed Rest, Spaceflight, Mars simulation
Cell Lines	Radiation (Ground), Simulated uG, Spaceflight, Parabolic Flight



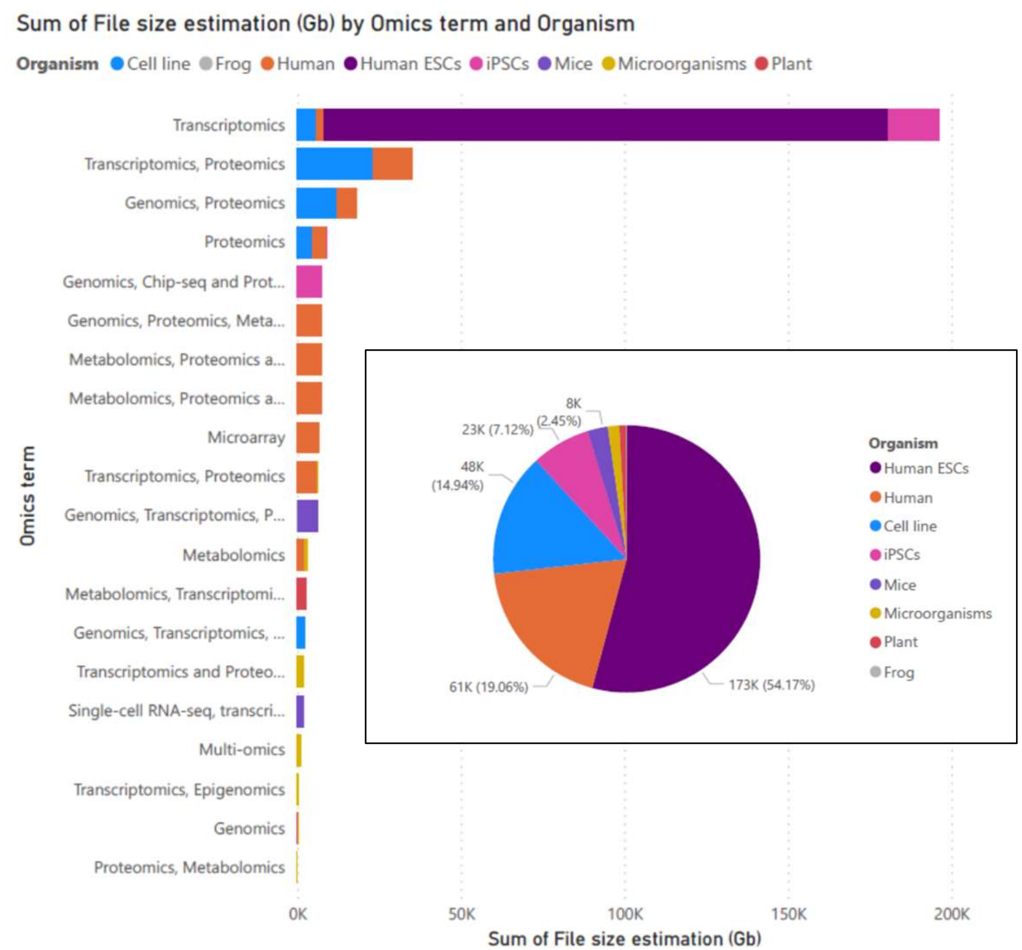
*Data originates from Gene Expression Omnibus (GEO) and EMBL's European Bioinformatics Institute (EMBL-EBI) repositories.

European Space Agency data arriving in OSDR: ~318 TB

Total of 39 Omics experiments, of which 21 are Multi-omics



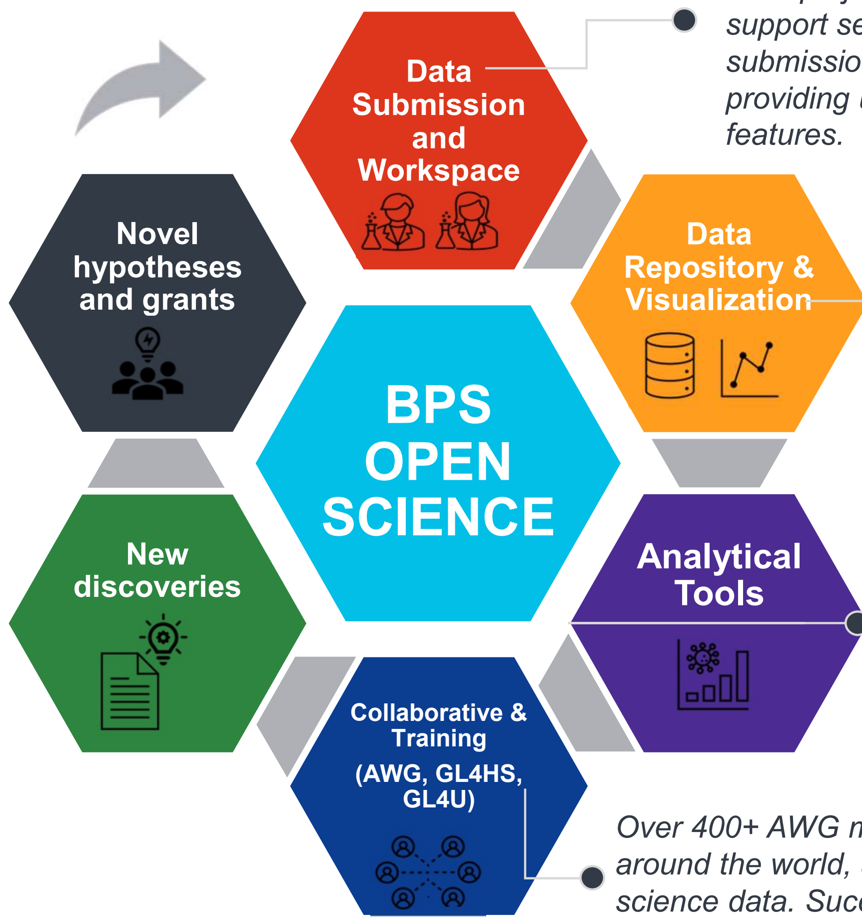
Platform	Sum of File size estimation (Gb)
Bed Rest	48800
DI	12000
IBPER	203230
ISS	54866
Parabolic flight	100
Total	**318996**



Open Science Data Repository ecosystem

Multi-project web-based submission portal to support self-service metadata curation and data submission within FAIR Guidelines. Workspace providing users file storage in S3 and sharing features.

Open access data enables discovery of new hypotheses and new ideas for grant proposal. Data from those new research experiments are generated and deposited back into GeneLab.



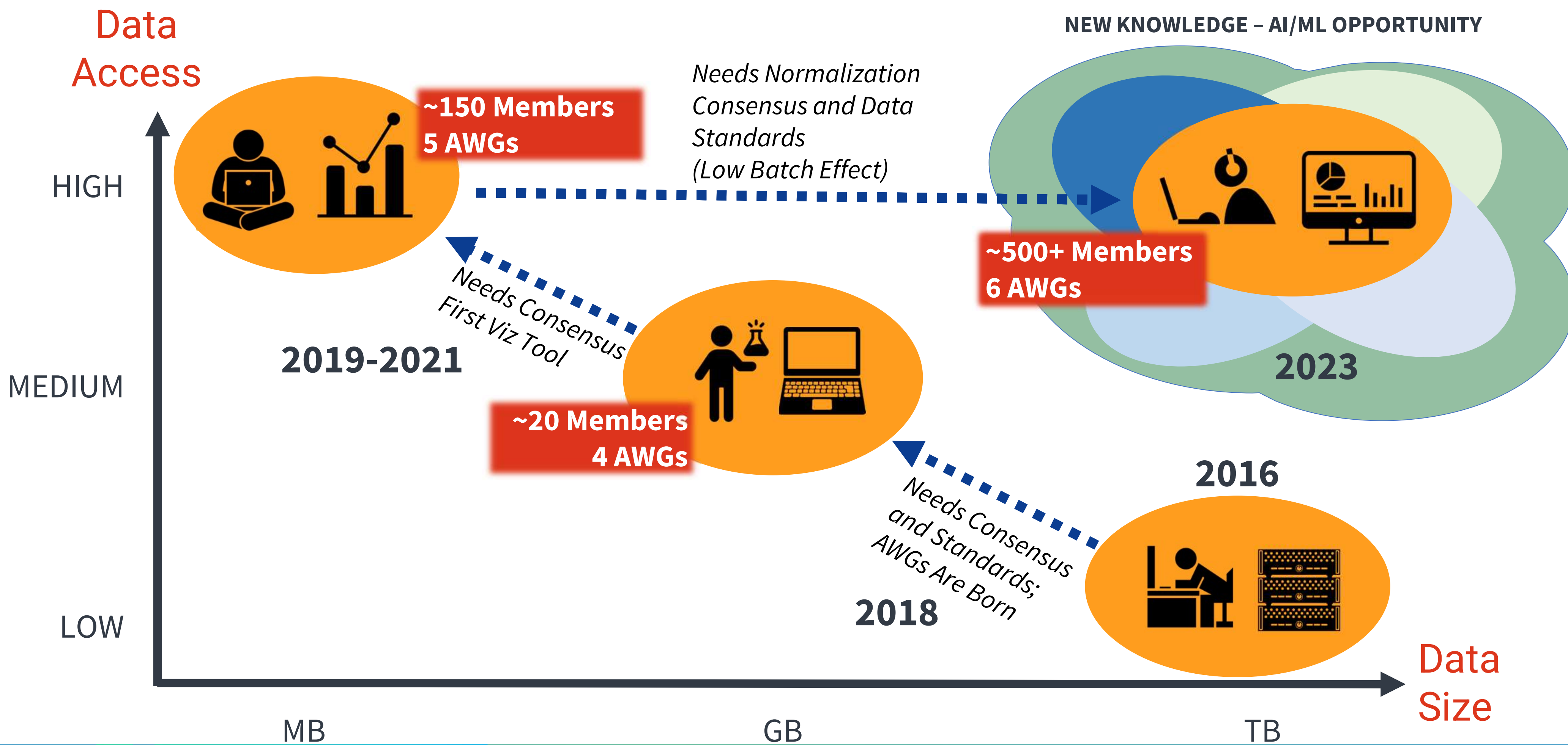
Open access and well curated data repositories including data from various model organisms from microbes to plants to humans covering the full range of biological assay technologies. Interactive visualization portal to enable knowledge discovery.

Enabling new publications and research through data reuse and performing meta-analysis.

Toolshed of bioinformatics tools with user-friendly interface for data analysis at any level (students, citizen scientists, PIs).

Over 400+ AWG members from around the world, analyzing open science data. Successful training programs developed for high school and university students.

A Brief History of the Open Science AWGs



Open Science Analysis Working Groups

ANIMAL

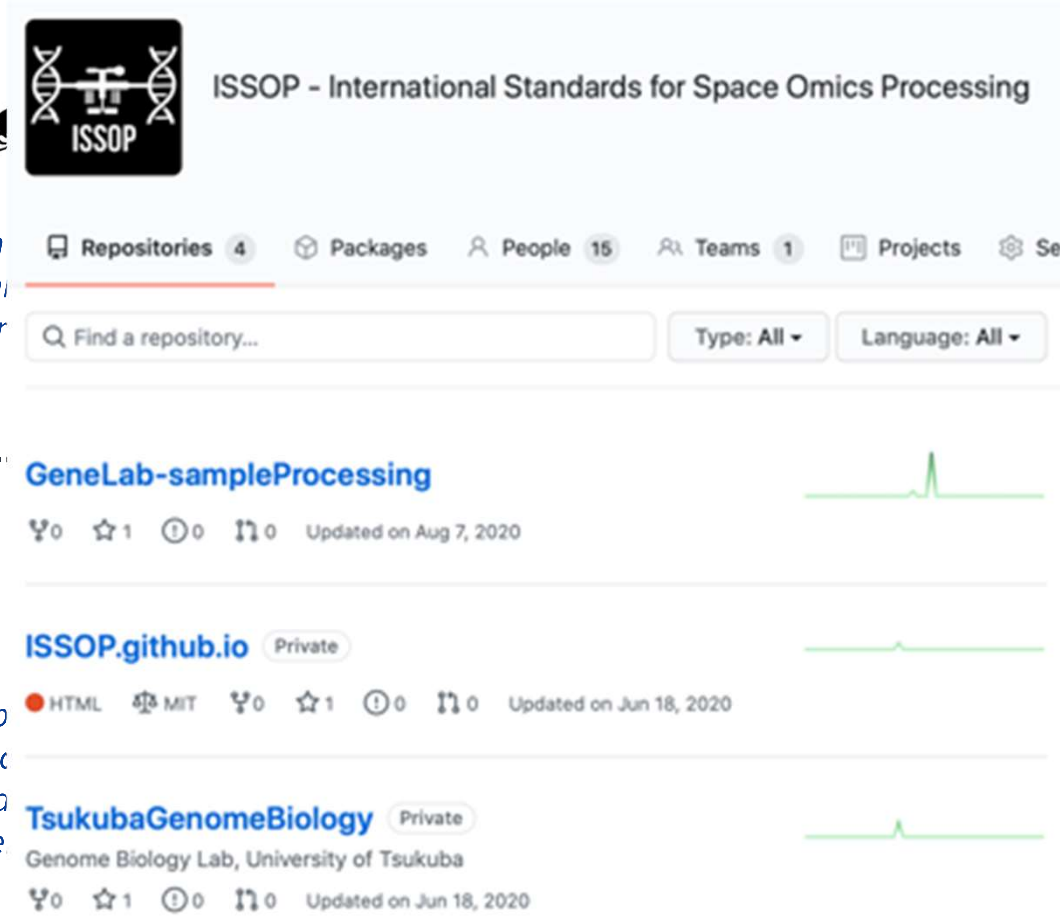
104 members

Facilitates the use of omics in understanding basic mechani which animals and constituer tissues and cells adapt to the spaceflight environment.

PLANTS

88 members

Share and discuss the latest developments in Astrob – the discipline of botany conc with interactions between pla biology and space environme



ROBES

nbers

s on analyzing microbial datasets GeneLab that includes gene- sion, proteomic, metabolomic and mental metagenomic datasets.

SDA

embers

ck on science data and metadata rds for physiological, phenotypic, and oral datasets to be reusable. Datasets om raw to processed-results data, and tabular, bioimaging, and video s.

*Consist of **500+ scientists** from multiple space agencies, international institutions, and industry. Scientists meet monthly with each group to provide feedback, develop standards, and analyze data.*

We invite you to join - https://genelab.nasa.gov/awg/join!

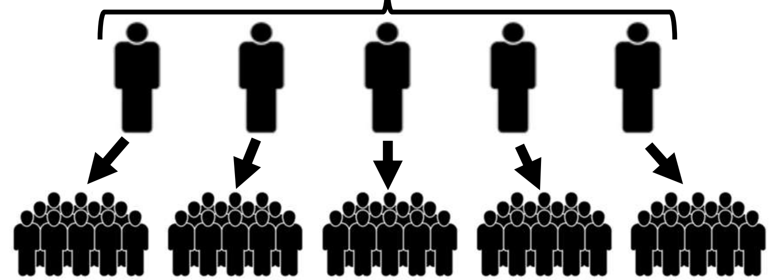
Bootcamps—Enabling our scientific community

GL4U Direct Approach
MSI/HBCU Students

2021 – Space biology & RNAseq
- Virtual 1-week long bootcamp with 14 **SJSU** students
- Collaboration with USRA/SJSU; SJSU compute system
- Space biology- and RNAseq-specific lectures and hands-on instruction using Jupyter Notebooks (JNs)

2023 – Space biology & Amplicon Seq
- In-person 4-day bootcamp with 20-40 **CSULA** students
- Collaboration with JPL/CSULA; NSF ACCESS compute system
- Space biology- and Amplicon Seq-specific lectures and JNs

GL4U Indirect Approach
MSI/HBCU Educators

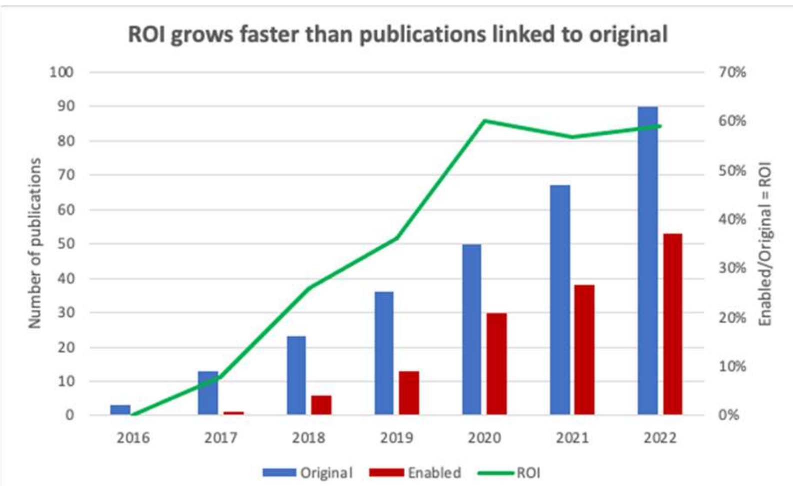


2022 – Space biology & RNAseq
- Virtual ~1.5-week long bootcamp
- Collaboration with JPL; SMCE compute system
- 6 professors and 4 graduate students from **4 HBCUs/MSIs**
- Space biology- and RNAseq-specific lectures and hands-on instruction using JNs
- Resources to teach at home institution

Metrics for success

90
Original Publications linked to OSDR (GeneLab/ALSDA)

52
Publications Enabled by OSDR System, Data Sharing

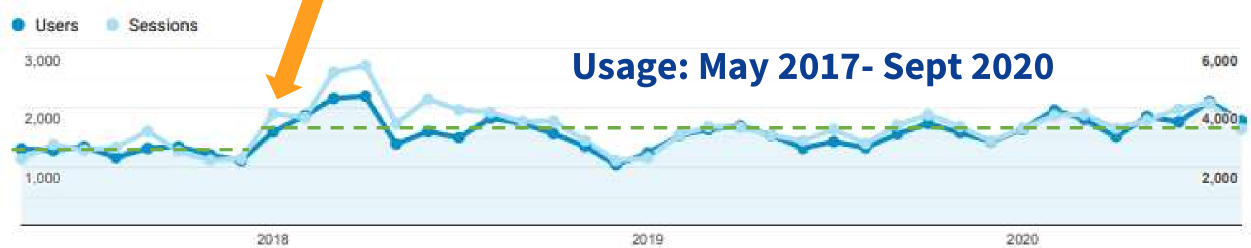


***52** publications (**13 publications produced by the AWGs**) using data available in OSDR.*

Introduction of Analysis Working Groups

3327
Avg Monthly Sessions

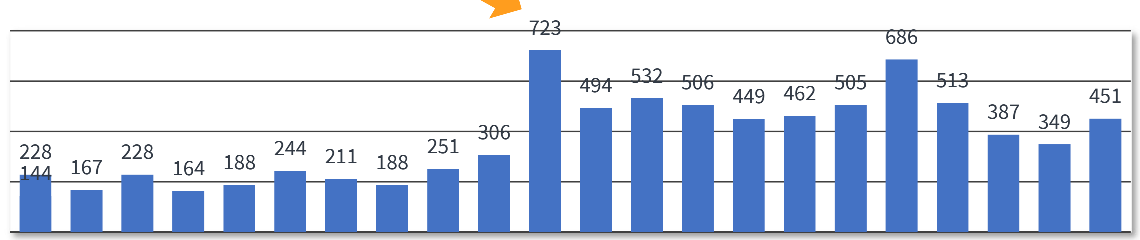
1558
Avg Monthly Users



Introduction of Processed Data and Visualization

360
Avg Monthly Downloads

7189
Total Study Downloads

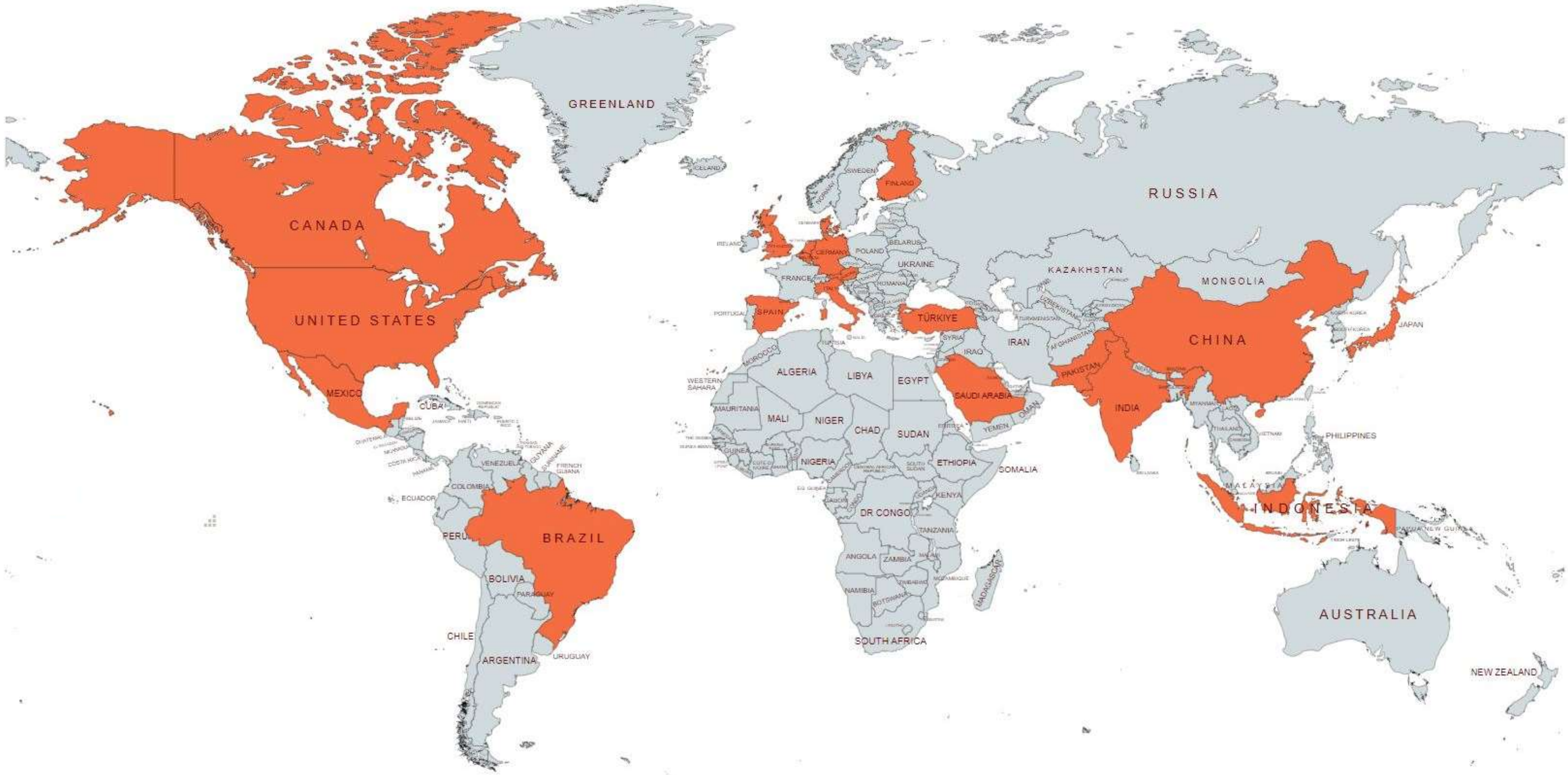


Nov 2019

Making data accessible worldwide

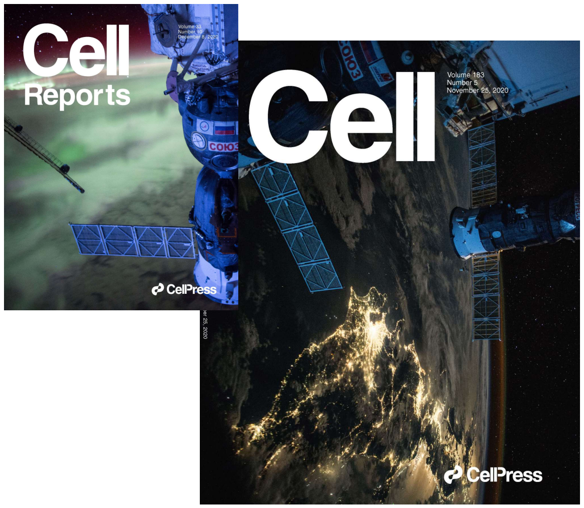
Top 20 Countries Utilizing GeneLab:

- US
- China
- UK
- India
- Germany
- Spain
- Canada
- Brazil
- Japan
- Netherlands
- Turkey
- Belgium
- Indonesia
- Italy
- Mexico
- Pakistan
- Saudi Arabia
- Austria
- Denmark
- Finland

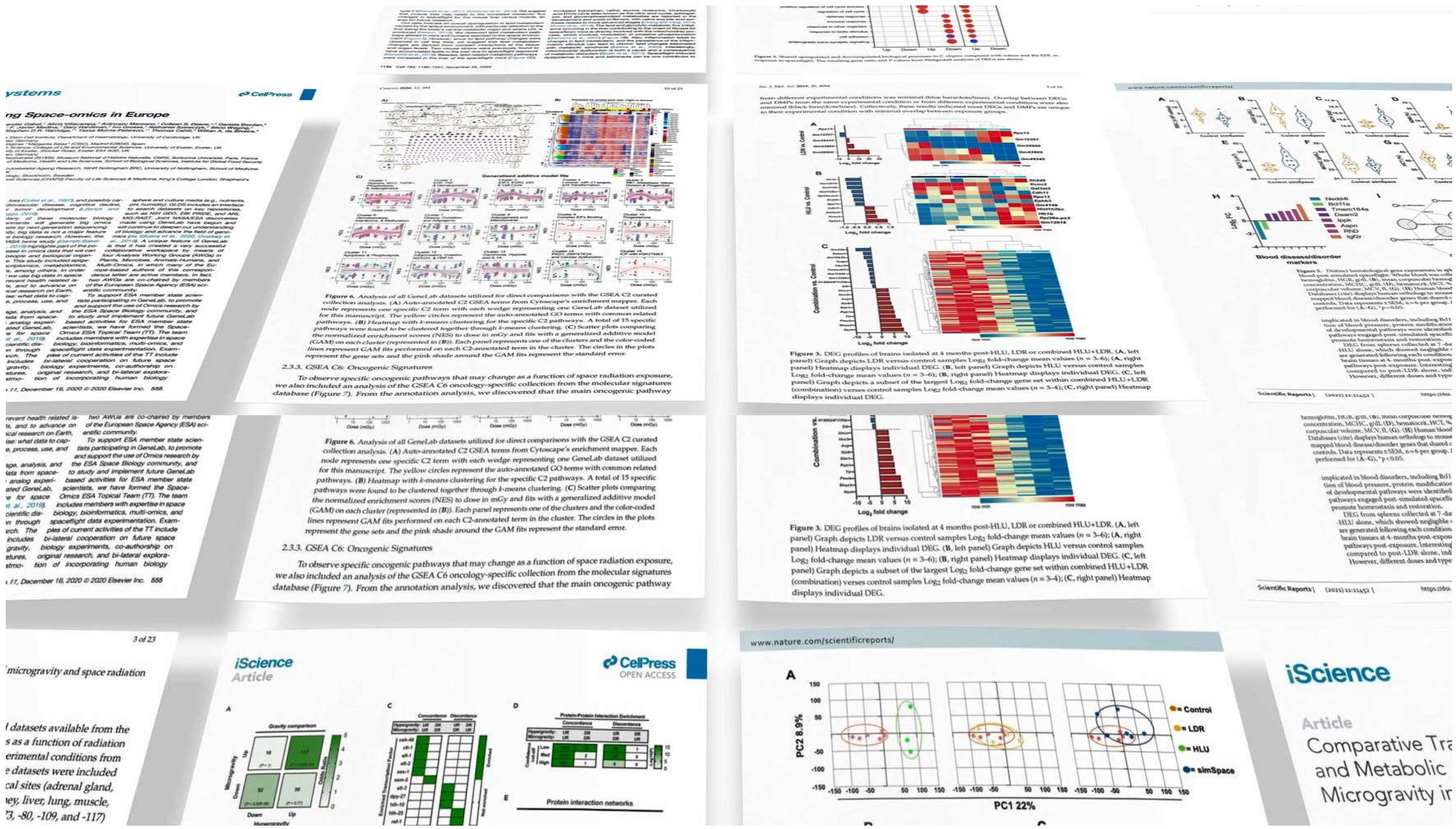


Re-use of Data and Enabling New Discoveries

29 publications (10 publications produced by the AWGs) using data available in GeneLab.



A coordinated package of **29 scientific papers** published in **five Cell Press journals** featuring **9 papers** utilizing data or resources in GeneLab. *High impact research highlights.*



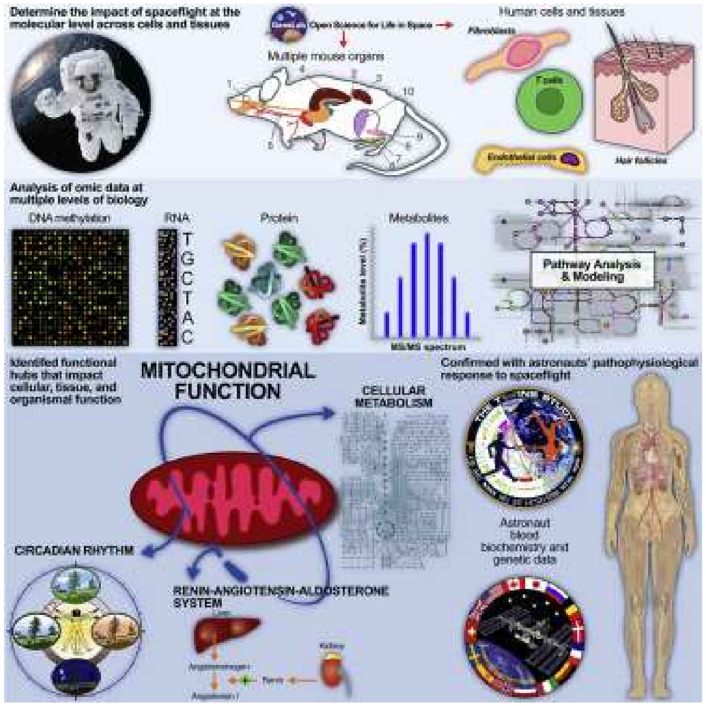
Comprehensive Multi-omics Analysis Reveals Mitochondrial Stress as a Central Biological Hub for Spaceflight Impact

Authors: Willian A. da Silveira, Hossein Fazelinia, Sara Brin Rosenthal, Evagelia C. Laiakis, Man S Kim, Cem Meydan, Yared Kidane, Komal S. Rathi, Scott M. Smith, Benjamin Stear, Yue Ying, Yuanchao Zhang, Jonathan Foox, Susana Zanello, Brian Crucian, Dong Wang, Adrienne Nugent, Helio A. Costa, Sara R. Zwart, Sonja Schrepfer, R. A. Leo Elworth, Nicolae Sapoval, Todd Treangen, Matthew MacKay, Nandan S. Gokhale, Stacy M. Horner, Larry N. Singh, Douglas C. Wallace, Jeffrey S. Willey, Jonathan C. Schisler, Robert Meller, J. Tyson McDonald, Kathleen M. Fisch, Gary Hardiman, Deanne Taylor, Christopher E. Mason, Sylvain V. Costes, Afshin Beheshti

Journal: Cell

Highlights:

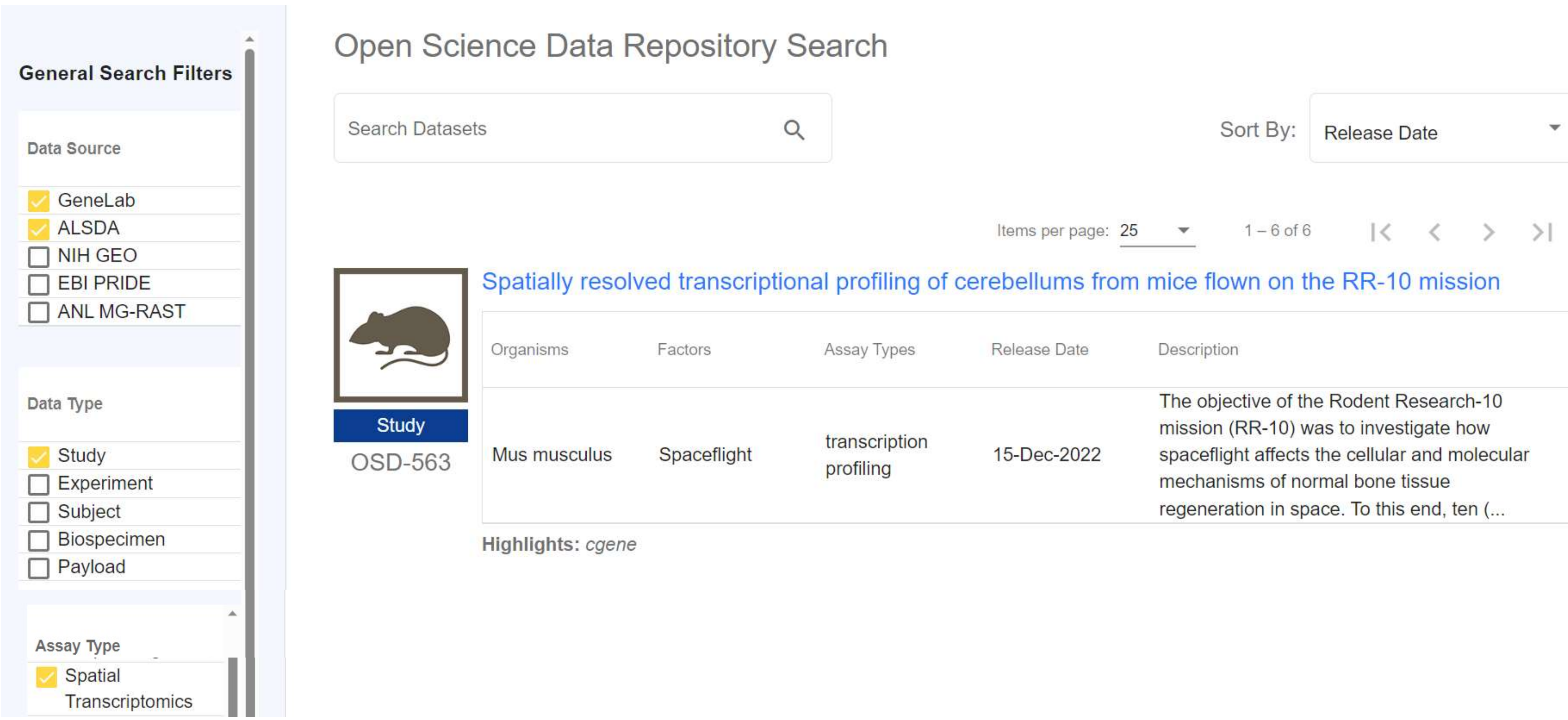
- Multi-omics analysis and techniques with NASA's GeneLab platform.
- The largest cohort of astronaut data to date utilized for analysis.
- Mitochondrial dysregulation driving spaceflight health risks.
- NASA Twins Study data validates mitochondrial dysfunction during space missions.



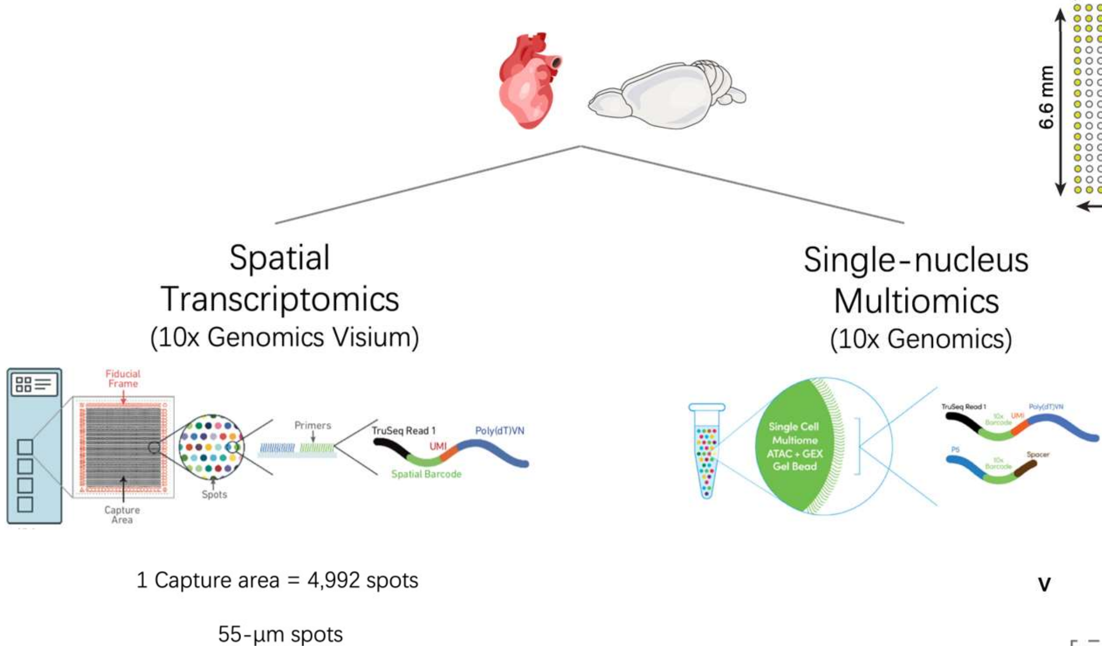
Relevance and Impact:

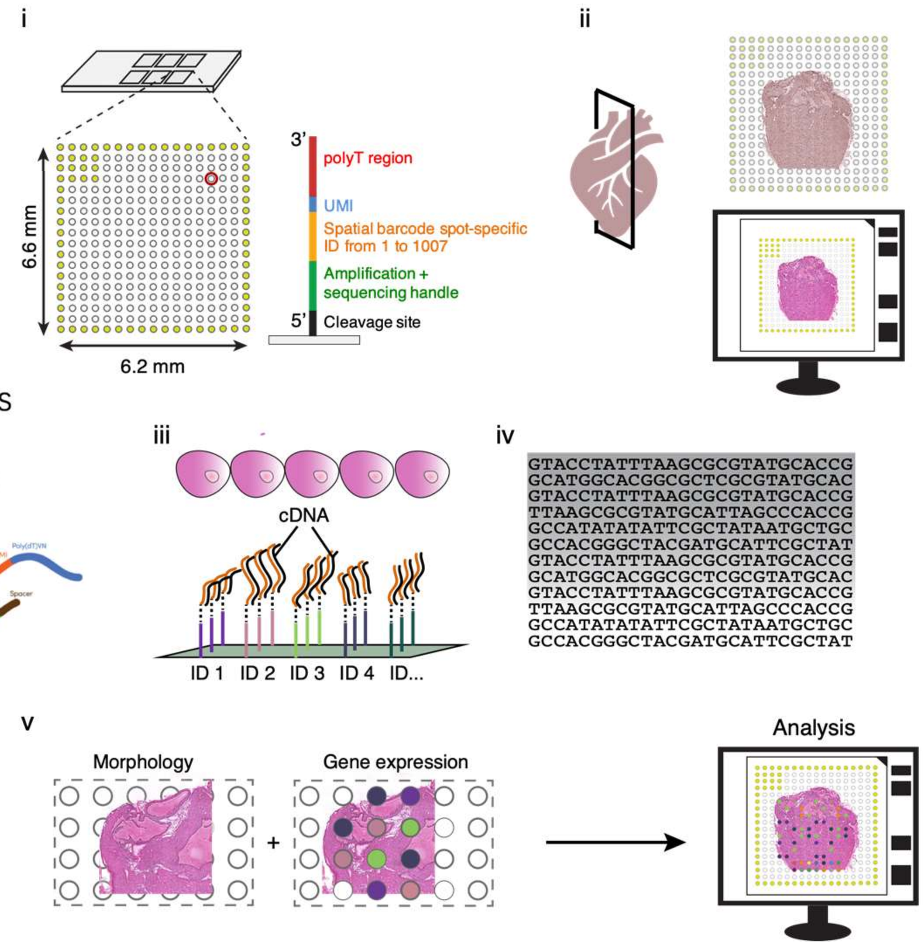
- Uncovered insights into fundamental biological mechanisms affected by spaceflight.
- Power of comparing and integrating multiple omics and data types to understand further how life adapts to spaceflight conditions.
- This concept can guide new nutritional and pharmaceutical interventions and studies that will increase the viability of long-term human-crewed space missions.

Spatial Transcriptomics for Space Samples - 10x Genomics Visium and Workflow

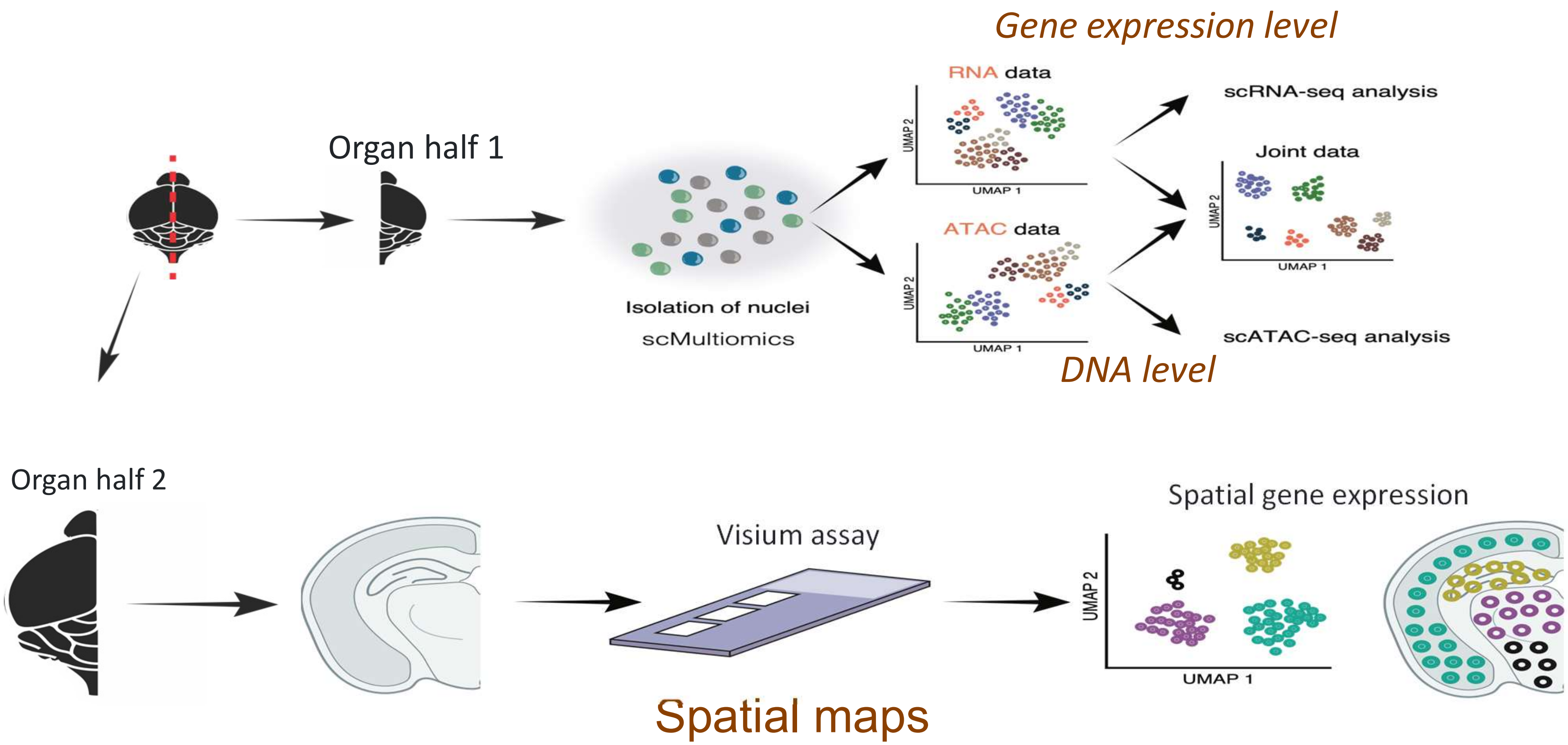


Spatial Transcriptomics for Space Samples - 10x Genomics Visium and Workflow



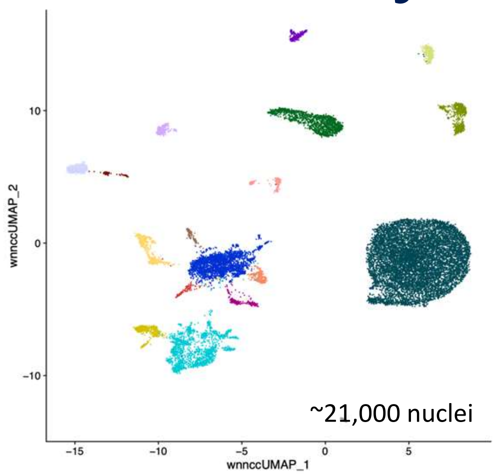


Multi-scale / Multi-modality (DNA, Genes, Tissue)



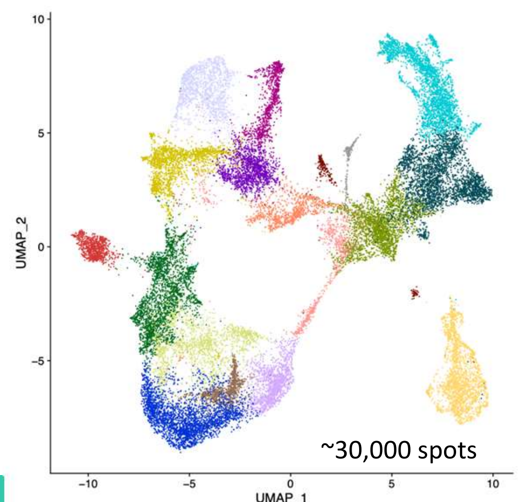
Brain - Preliminary

Multiomics

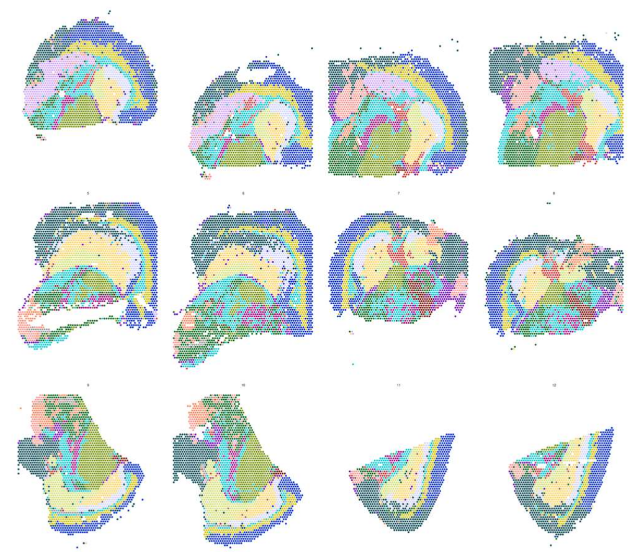


- 0 – Neurogenesis I
- 1 – Synaptic transmission I
- 2 – Neuronal activity, synaptic transmission I
- 3 – Myelination
- 4 – Astrocytes
- 5 – Glutamatergic synaptic transmission I
- 6 – Glutamatergic synaptic transmission II
- 7 – Glial development
- 8 – Glutamatergic synaptic transmission III
- 9 – Neurogenesis II
- 10 – Microglia
- 11 – GABAergic synaptic transmission
- 12 – Neurovasculature
- 13 – Neuronal activity, synaptic transmission II
- 14 – Neuroendocrine activity
- 15 – Neuronal activity, synaptic transmission III
- 16 – Glutamatergic synaptic transmission IV
- 17 – Synaptic transmission II

Visium



- 0 – Midbrain neurons
- 1 – Cortical neurons, top layer
- 2 – Neuroendocrine neurons
- 3 – Amygdalar and piriform neurons
- 4 – White matter
- 5 – Thalamic neurons
- 6 – Hippocampal CA1 neurons, pyramidal layer
- 7 – Auditory and entorhinal neurons
- 8 – Hippocampal CA3 neurons
- 9 – Cortical neurons, bottom layer
- 10 – Hippocampal CA1 neurons, striatam radiatum
- 11 – Dendate gyrus neurons
- 12 – Corpus callosum
- 13 – Neurovasculature
- 14 – Caudate/putamen neurons
- 15 – Retrosplenial neurons
- 16 – Choroid plexus and subventricular structures
- 17 – Choroid plexus



Publications

Cell Reports Methods

Volume 2, Issue 11, 21 November 2022, 100325

Perspective

Challenges and considerations for single-cell and spatially resolved transcriptomics sample collection during spaceflight

Eliah G. Overbey [1 2 18], Saswati Das [3 18], Henry Cope [4], Pedro Madrigal [5], Zaneta Andrusivova [6], Solène Frapard [6], Rebecca Klotz [7], Daniela Bezdan [8 9 10], Anjali Gupta [11], Ryan T. Scott [7], Jiwoon Park [1], Dawn Chirko [1], Jonathan M. Galazka [12], Sylvain V. Costes [12], Christopher E. Mason [1 2 13 14], Raul Herranz [15], Nathaniel J. Szewczyk [4 16], Joseph Borg [17], Stefania Giacomello [6]

Nature Communications - Under Review

Spatially resolved multiomics on the neuronal effects induced by spaceflight

Yuvarani Masarapu[1#], Egle Cekanaviciute[2#], Zaneta Andrusivova[1#], Jakub O. Westholm[3], Åsa Björklund[4], Robin Fallegger[5], Pau Badia-i-Mompel[5], Valery Boyko[2], Shubha Vasisht[6], Amanda Saravia-Butler[2], Samrawit Gebre[2], Enikö Lázár[7,8], Olaf Bergmann[7], Deanne M Taylor[6,9], Douglas C. Wallace[9], Christer Sylven[10], Julio Saez-Rodriguez[5], Jonathan M. Galazka[2*], Sylvain V. Costes[2*], Stefania Giacomello[1*]

NASA Decadal Survey

TOPICAL: TAKING SPACE BIOLOGY TO THE NEXT ERA BY ENSURING STANDARDIZATION, APPLICABILITY OF *IN SITU* TISSUE ANALYSIS AND COMPUTATIONAL PREDICTIONS

Graphical abstract

- Standardization of omics experiments is required
- Single-cell RNA-sequencing and spatially-resolved transcriptomics experiments are needed to extract more information from biological specimens exposed to spaceflight
- Artificial Intelligence will reduce the time required by astronauts to physically work on scientific experiments

Primary author (ISSOP* Science co-chair):
Stefania Giacomello, Science for Life Laboratory, KTH Royal Institute of Technology, Sweden
Phone number: +46 (0)73 6697 828
Email address: stefania.giacomello@scilifelab.se

Co-authors (ISSOP* members):
Rebecca Klotz, KBR/NASA Ames Research Center, Moffett Field, California, U.S.A.
Ryan T. Scott, KBR/NASA Ames Research Center, Moffett Field, California, U.S.A.
Saswati Das, Central Government Health Services, Dr Ram Manohar Lohia Hospital and Atal Bihari Vajpayee Institute of Medical Sciences, India

Co-authors (Space Omics TT# members):
Joseph Borg, Faculty of Health Sciences, University of Malta, Malta
Pedro Madrigal, European Molecular Biology Laboratory, European Bioinformatics Institute, EMBL-EBI, United Kingdom

Endorsed by (ISSOP* members):
Amber M. Paul, Embry-Riddle Aeronautical University, NASA Ames Research Center, U.S.A.
Deanne M. Taylor, University of Pennsylvania, Perelman Medical School, and The Children's Hospital of Philadelphia, U.S.A.
Henry Cope, School of Medicine, University of Nottingham, United Kingdom
Jonathan M. Galazka, Space Biosciences Division, NASA Ames Research Center, U.S.A.
Manuel A. Fernandez-Rojo, IMDEA-Food Institute, Hepatic Regenerative Medicine Group, Spain
Masafumi Muratani, University of Tsukuba, Japan
Solène Frapard, Science for Life Laboratory, KTH Royal Institute of Technology, Sweden
Sunny Narayanan, NASA Post-Doctoral Fellow, Florida State University, U.S.A.
Sylvain V. Costes, Space Biosciences Division, NASA Ames Research Center, U.S.A.

*International Standards for Space Omics Processing
#Space Omics Topical Team (funded by ESA – European Space Agency)

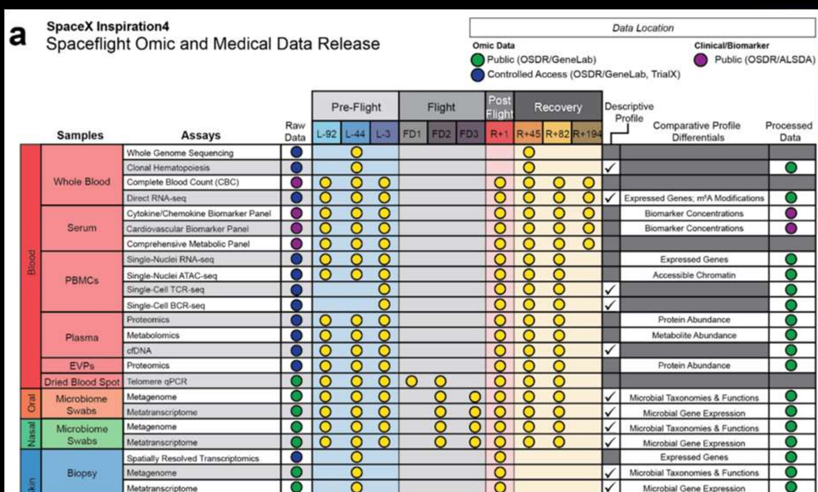
Stefania Giacomello

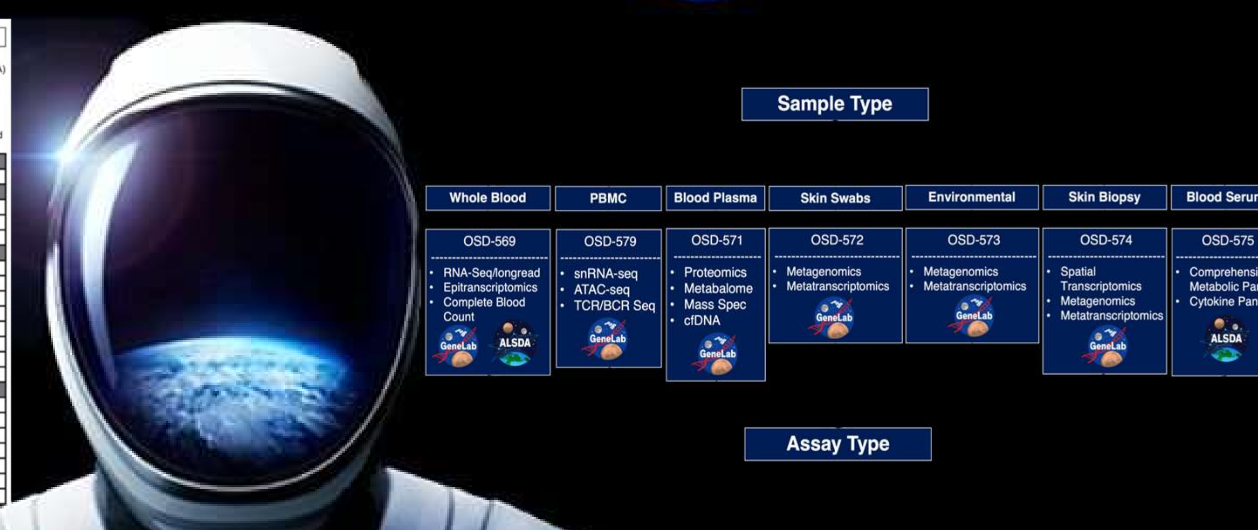
SciLifeLab

@giacomello_lab

Engaging with commercial entities: I4 Human Data into OSDR

INSPIRATI4N





- Spring '23 Nature 'I4/JAXA/SB Package'
- Credit: Chris Mason's Lab; Overbey et al., 2023
- In Peer Review at *Nature*

OSDR able to ingest and properly curate all data for reuse

3 New ALSDA Assay Metadata 'Configs'

Comprehensive Metabolic Panel (Spectrophotometry & Ion Selective Electrode) | Complete Blood Count (Hematology Analyzer) | Cytokine Panel (Multiplex)

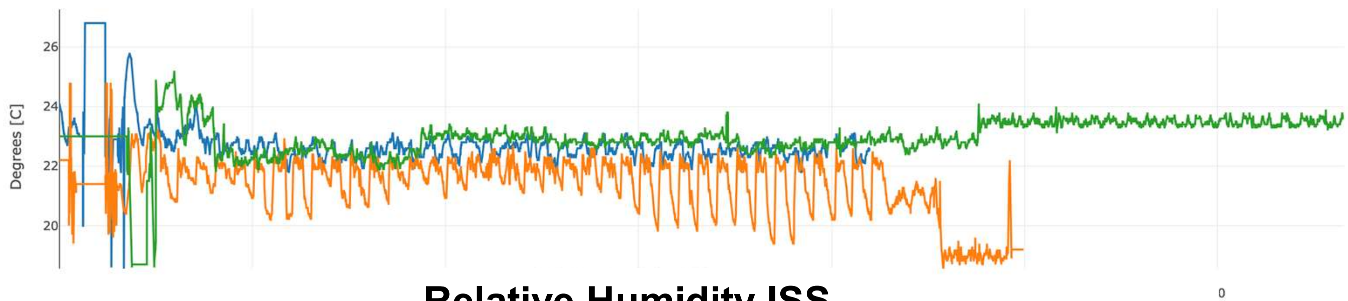
Credit for slide design/content: Kristen Peach, GeneLab Curator

Environmental Data App

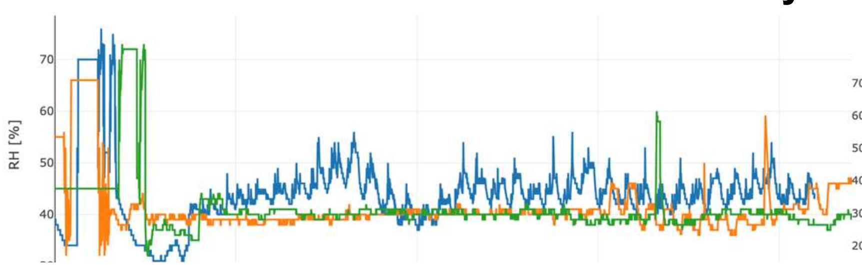
Spaceflight Mission Comparison

RR1
RR3
RR5

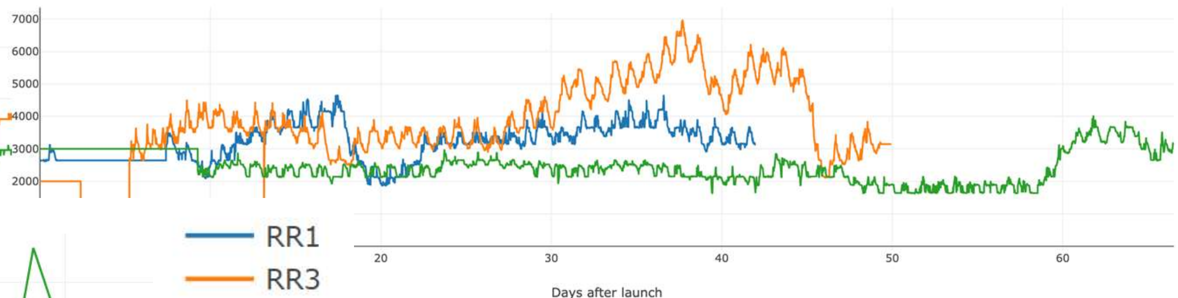
Temperature ISS (ºC)



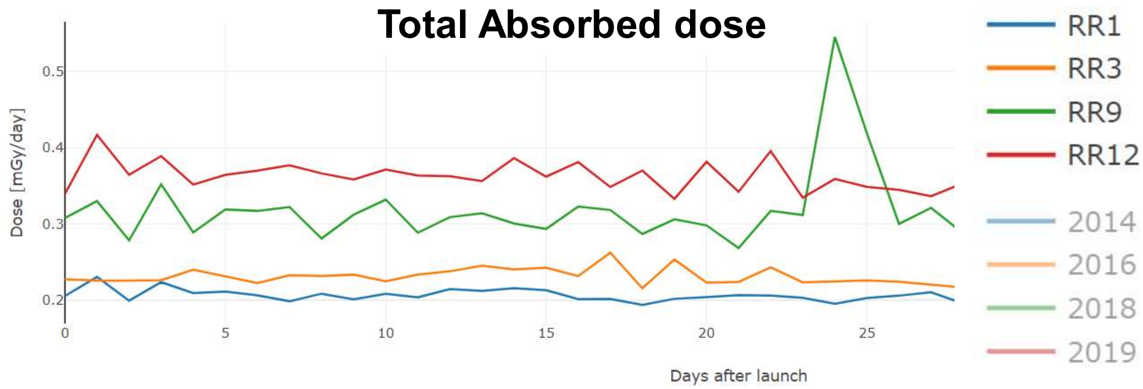
Relative Humidity ISS

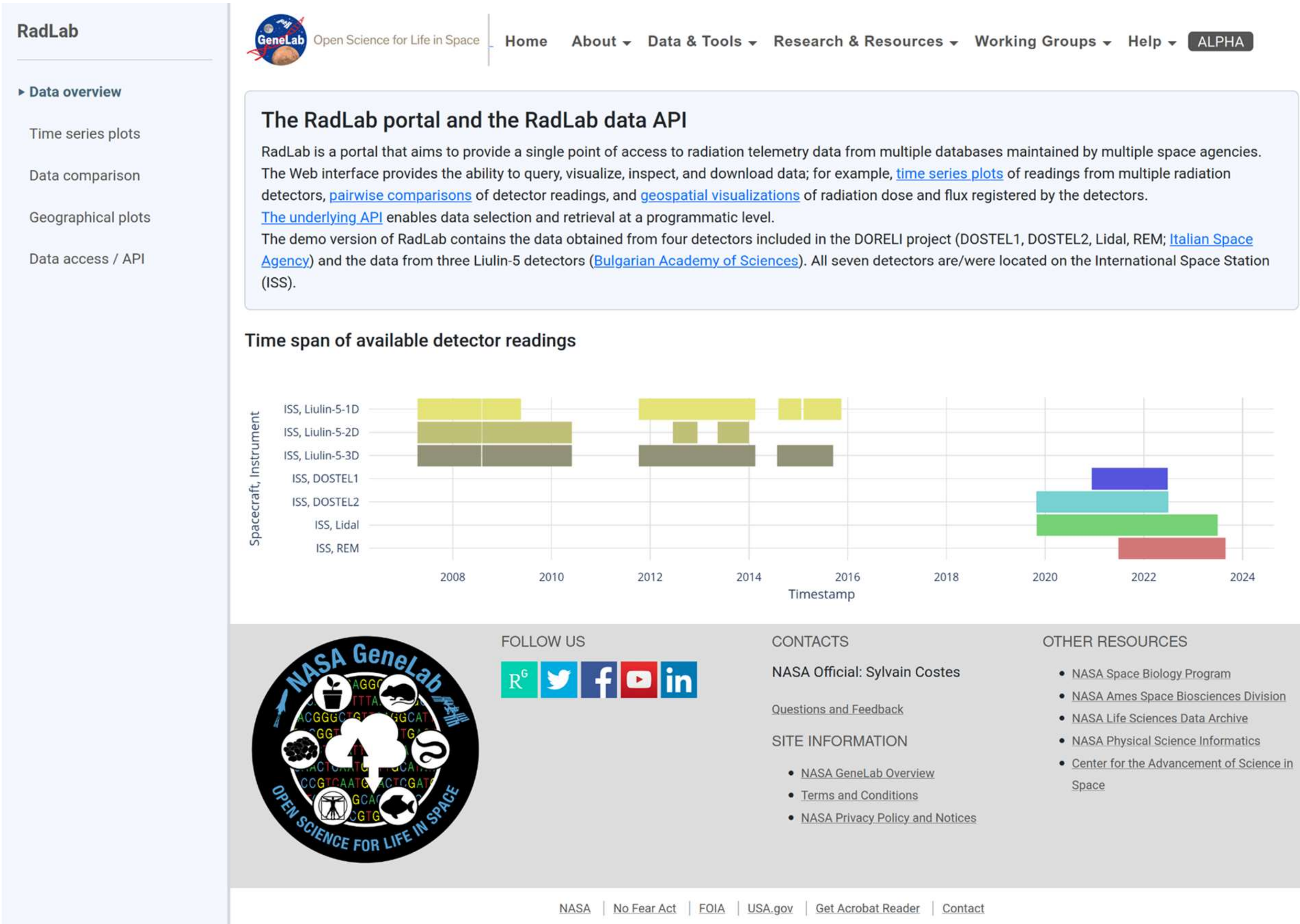


CO2 ISS

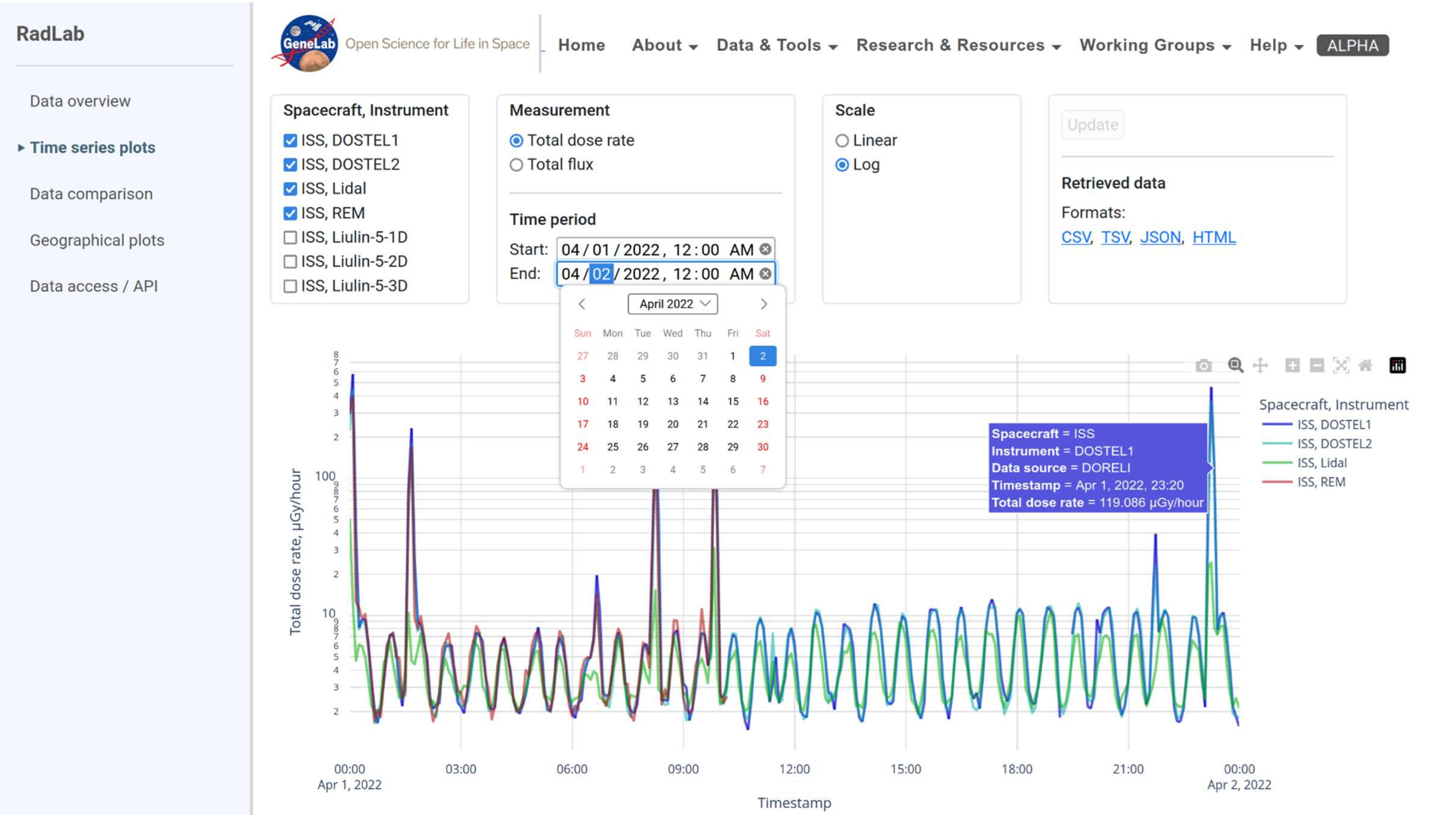


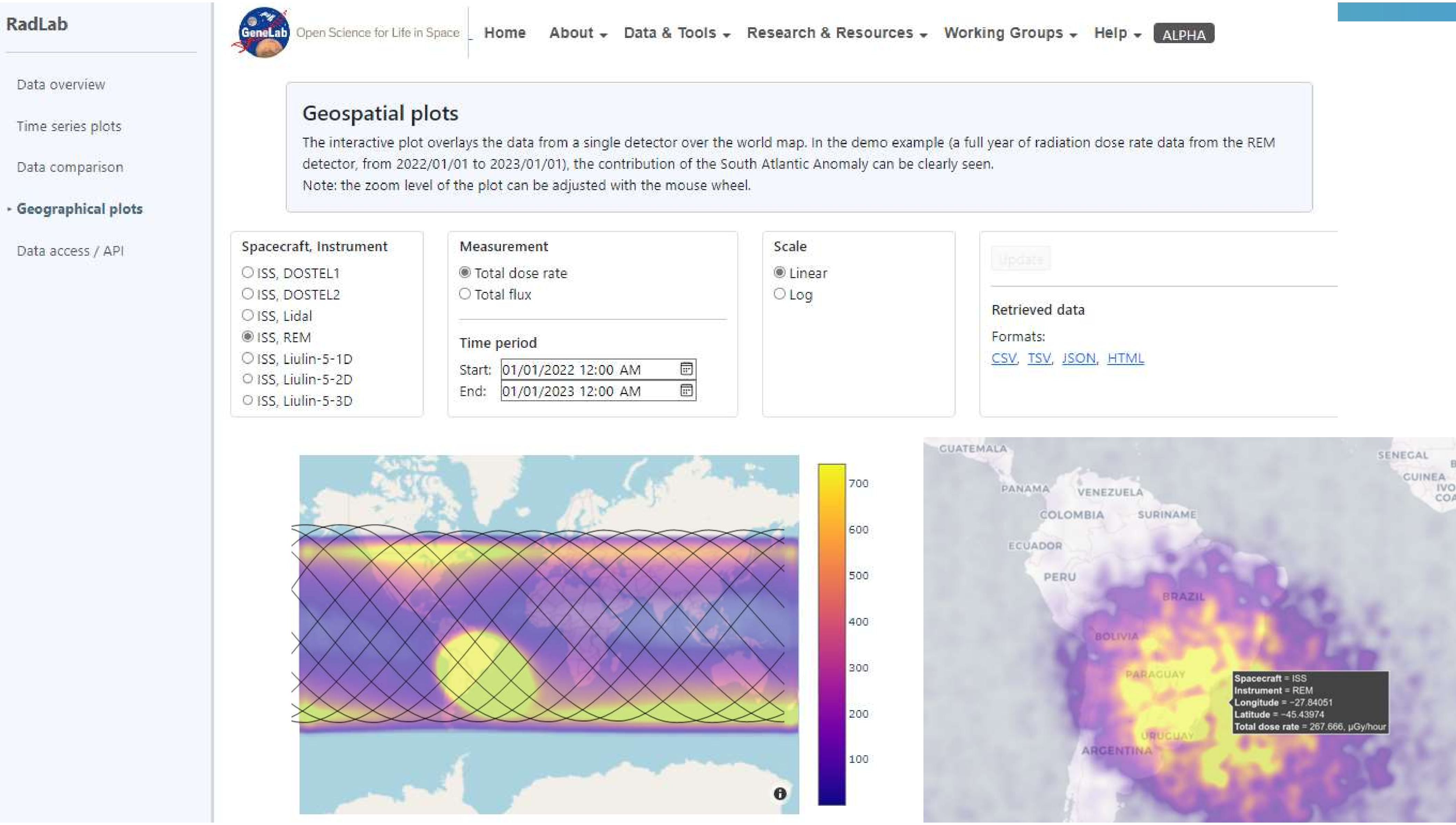
Total Absorbed dose





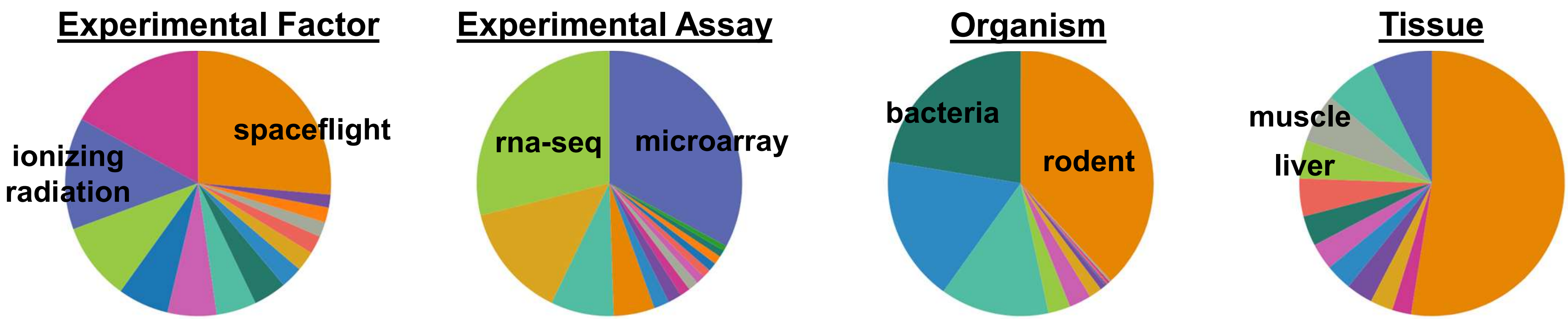
RadLab
Data overview
Time series plots
Data comparison
Geographical plots
Data access / API
GeneLab
Open Science for Life in Space
Home
About
Data & Tools
Research & Resources
Working Groups
Help
ALPHA
The RadLab portal and the RadLab data API
RadLab is a portal that aims to provide a single point of access to radiation telemetry data from multiple databases maintained by multiple space agencies. The Web interface provides the ability to query, visualize, inspect, and download data; for example, time series plots of readings from multiple radiation detectors, pairwise comparisons of detector readings, and geospatial visualizations of radiation dose and flux registered by the detectors.
The underlying API enables data selection and retrieval at a programmatic level.
The demo version of RadLab contains the data obtained from four detectors included in the DORELI project (DOSTEL1, DOSTEL2, Lidal, REM; Italian Space Agency) and the data from three Liulin-5 detectors (Bulgarian Academy of Sciences). All seven detectors are/were located on the International Space Station (ISS).
Time span of available detector readings
Spacecraft, Instrument
ISS, Liulin-5-1D
ISS, Liulin-5-2D
ISS, Liulin-5-3D
ISS, DOSTEL1
ISS, DOSTEL2
ISS, Lidal
ISS, REM
2008
2010
2012
2014
2016
2018
2020
2022
2024
Timestamp
NASA GeneLab
OPEN SCIENCE FOR LIFE IN SPACE
FOLLOW US
CONTACTS
NASA Official: Sylvain Costes
Questions and Feedback
SITE INFORMATION
NASA GeneLab Overview
Terms and Conditions
NASA Privacy Policy and Notices
OTHER RESOURCES
NASA Space Biology Program
NASA Ames Space Biosciences Division
NASA Life Sciences Data Archive
NASA Physical Science Informatics
Center for the Advancement of Science in Space
NASA
No Fear Act
FOIA
USA.gov
Get Acrobat Reader
Contact


RadLab
Data overview
Time series plots
Data comparison
Geographical plots
Data access / API
GeneLab
Open Science for Life in Space
Home
About
Data & Tools
Research & Resources
Working Groups
Help
ALPHA
Spacecraft, Instrument
ISS, DOSTEL1
ISS, DOSTEL2
ISS, Lidal
ISS, REM
ISS, Liulin-5-1D
ISS, Liulin-5-2D
ISS, Liulin-5-3D
Measurement
Total dose rate
Total flux
Time period
Start: 04 / 01 / 2022 , 12 : 00 AM
End: 04 / 02 / 2022 , 12 : 00 AM
April 2022
Sun Mon Tue Wed Thu Fri Sat
Scale
Linear
Log
Update
Retrieved data
Formats:
CSV, TSV, JSON, HTML
Spacecraft = ISS
Instrument = DOSTEL1
Data source = DORELI
Timestamp = Apr 1, 2022, 23:20
Total dose rate = 119.086 µGy/hour
Spacecraft, Instrument
ISS, DOSTEL1
ISS, DOSTEL2
ISS, Lidal
ISS, REM
Total dose rate, µGy/hour
Timestamp
00:00 Apr 1, 2022
03:00
06:00
09:00
12:00
15:00
18:00
21:00
00:00 Apr 2, 2022


RadLab
Data overview
Time series plots
Data comparison
Geographical plots
Data access / API
GeneLab
Open Science for Life in Space
Home
About
Data & Tools
Research & Resources
Working Groups
Help
ALPHA
Geospatial plots
The interactive plot overlays the data from a single detector over the world map. In the demo example (a full year of radiation dose rate data from the REM detector, from 2022/01/01 to 2023/01/01), the contribution of the South Atlantic Anomaly can be clearly seen.
Note: the zoom level of the plot can be adjusted with the mouse wheel.
Spacecraft, Instrument
ISS, DOSTEL1
ISS, DOSTEL2
ISS, Lidal
ISS, REM
ISS, Liulin-5-1D
ISS, Liulin-5-2D
ISS, Liulin-5-3D
Measurement
Total dose rate
Total flux
Time period
Start: 01/01/2022 12:00 AM
End: 01/01/2023 12:00 AM
Scale
Linear
Log
Retrieved data
Formats:
CSV, TSV, JSON, HTML
700
600
500
400
300
200
100
GUATEMALA
SENEGAL
GUINEA
PANAMA
VENEZUELA
COLOMBIA
SURINAME
ECUADOR
PERU
BRAZIL
BOLIVIA
PARAGUAY
URUGUAY
ARGENTINA
Spacecraft = ISS
Instrument = REM
Longitude = −27.84051
Latitude = −45.43974
Total dose rate = 267.666, µGy/hour

Visualization and Analysis Platform

1. Filter Data



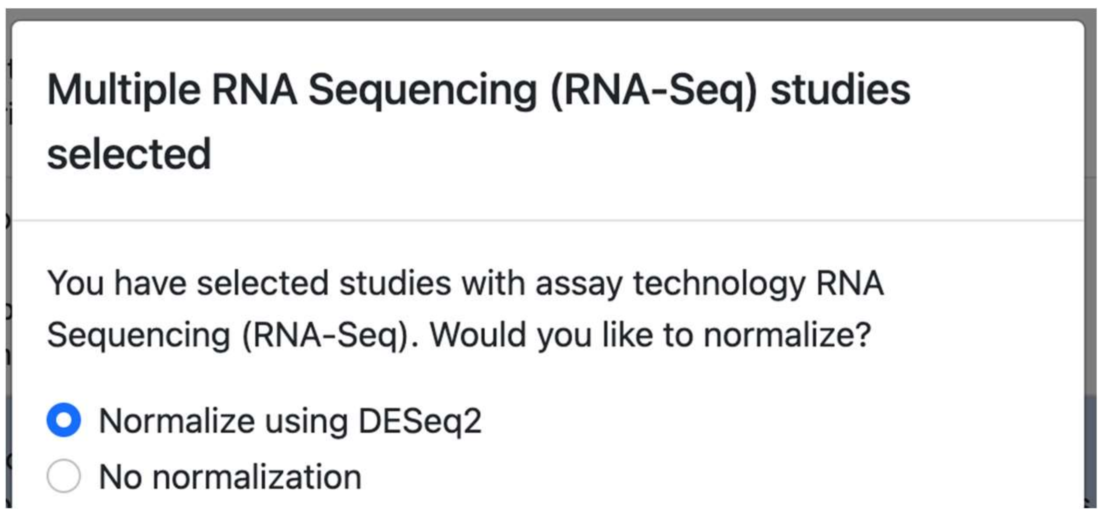
2. Select Datasets to Analyze

	OSD	Title	Assay	Organism	Tissue	Factor
☐	OSD-47	Rodent Research-1 (RR1) National Lab Validation Flight: Mouse liver transcriptomic, proteomic, epigenomic and histology data	RNA Sequencing (RNA-Seq)	Mus musculus	Liver	spaceflight,spaceflight.term accession number,spaceflight.term source ref,
☐	OSD-48	Rodent Research-1 (RR1) NASA Validation Flight: Mouse liver transcriptomic, proteomic, epigenomic and histology data	RNA Sequencing (RNA-Seq)	Mus musculus	Liver	dissection condition,spaceflight,spaceflight.term accession number,spaceflight.term source ref,

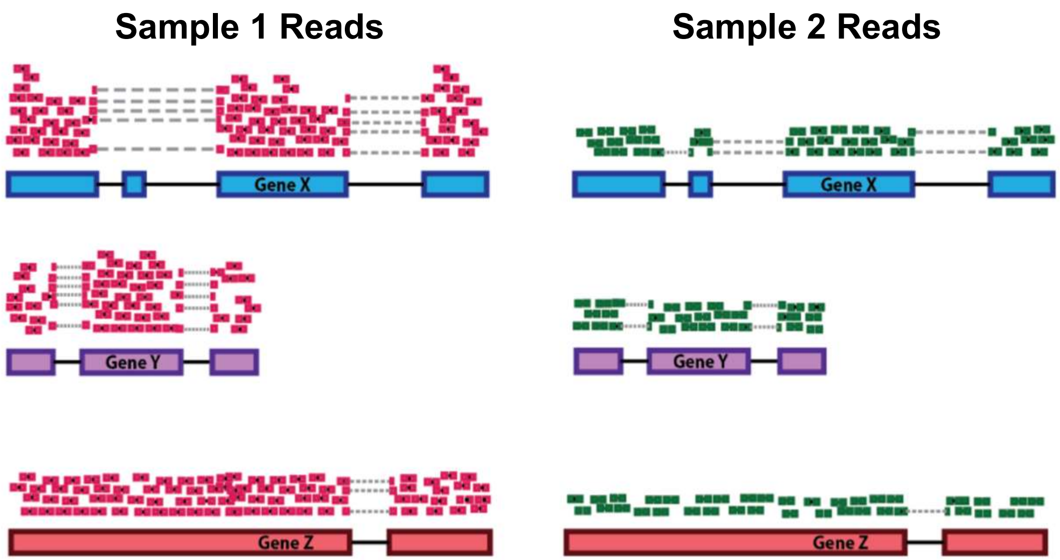
Visualize Study

Multi-Study Normalization

- Multiple RNA-seq studies can be *combined* and *normalized*
- DESeq2 “Median of Ratios” normalization
  - Ensures that all samples are directly comparable across experiments

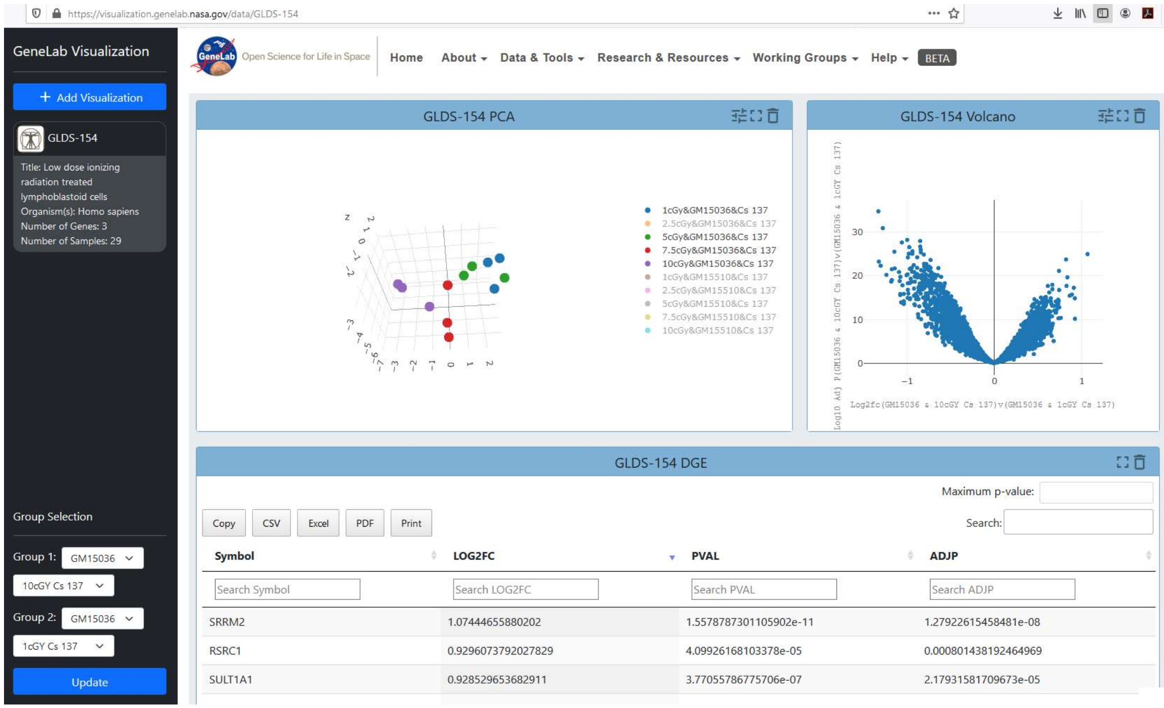


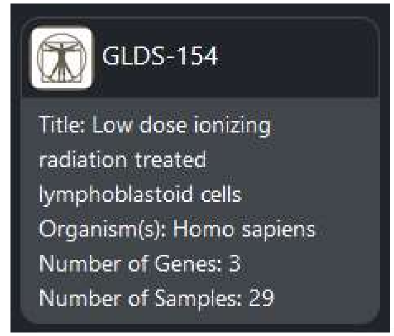
DESeq2 Median of Ratios

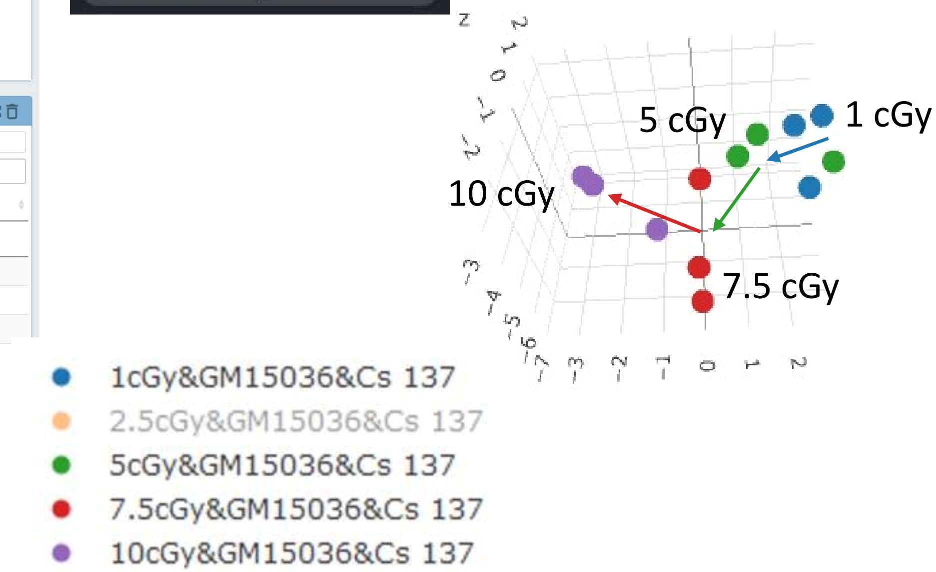


- Sample 1 was sequenced at higher depth than Sample 2
- Direct comparison without normalization makes it look like ***all*** genes are expressed more highly in Sample 1
- <u>DESeq2 Median of Ratios:</u>
  - Leverages the fact that most genes should be similar across samples
  - Borrows information across samples in order to calculate a “normalization factor” for each gene
  - Converts raw counts to normalized counts

Low dose radiation transcriptomic data can be visualized





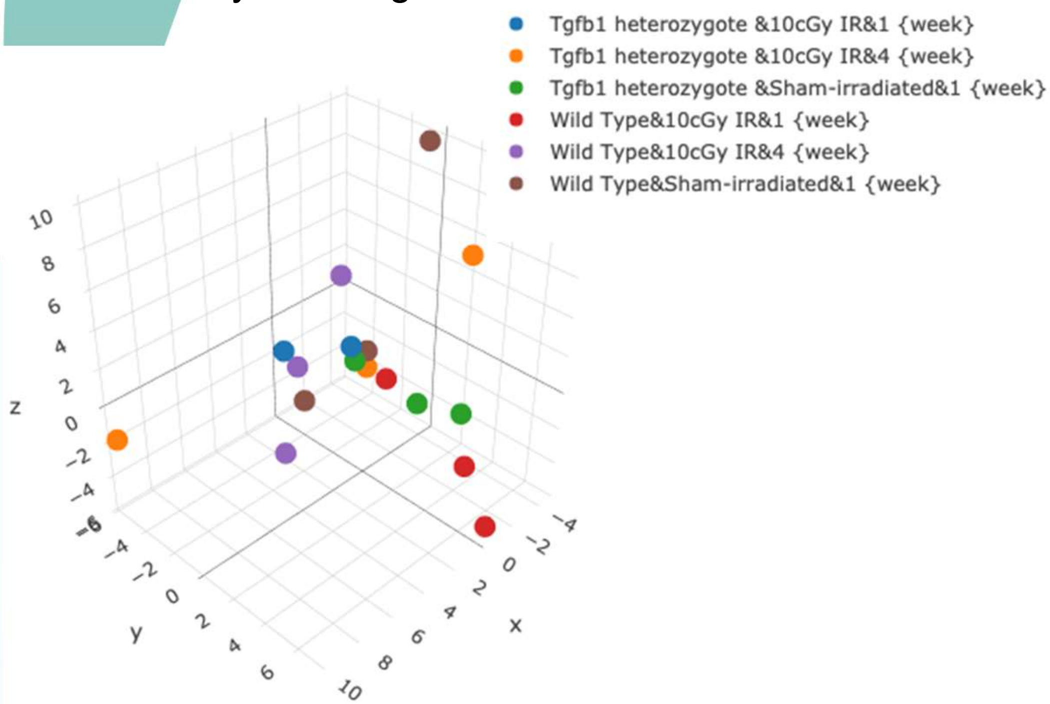


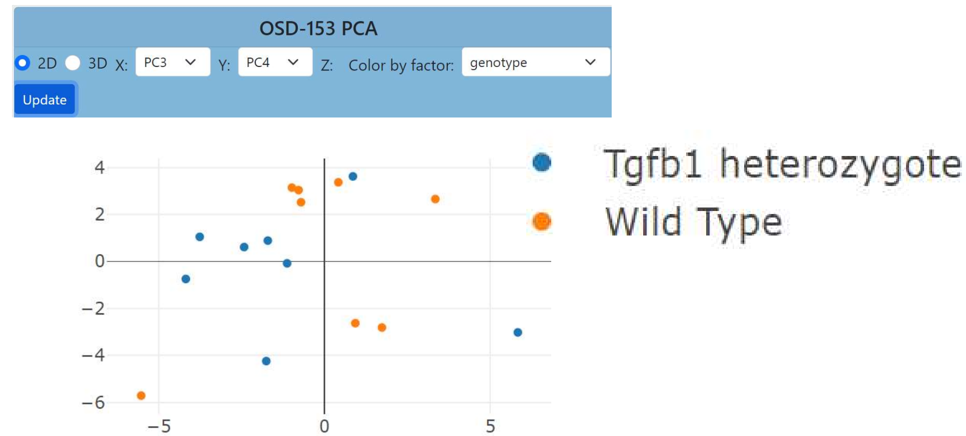
Wu P, Coleman M, Wyrobek AJ. "Low dose ionizing radiation treated lymphoblastoid cells", GeneLab, Version 3, http://doi.org/10.26030/hs0p-6w85

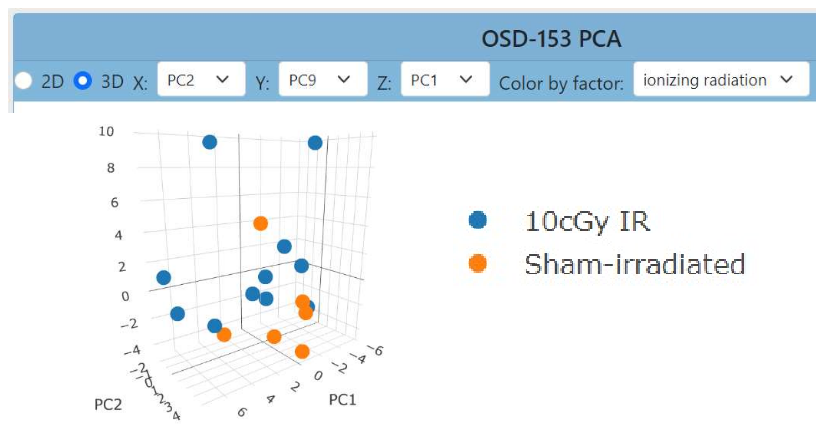
Low dose radiation transcriptomic data can be visualized

Visualization Platform

GLDS-153: Non-targeted effects of low dose ionizing radiation act via TGF-beta to promote mammary carcinogenesis



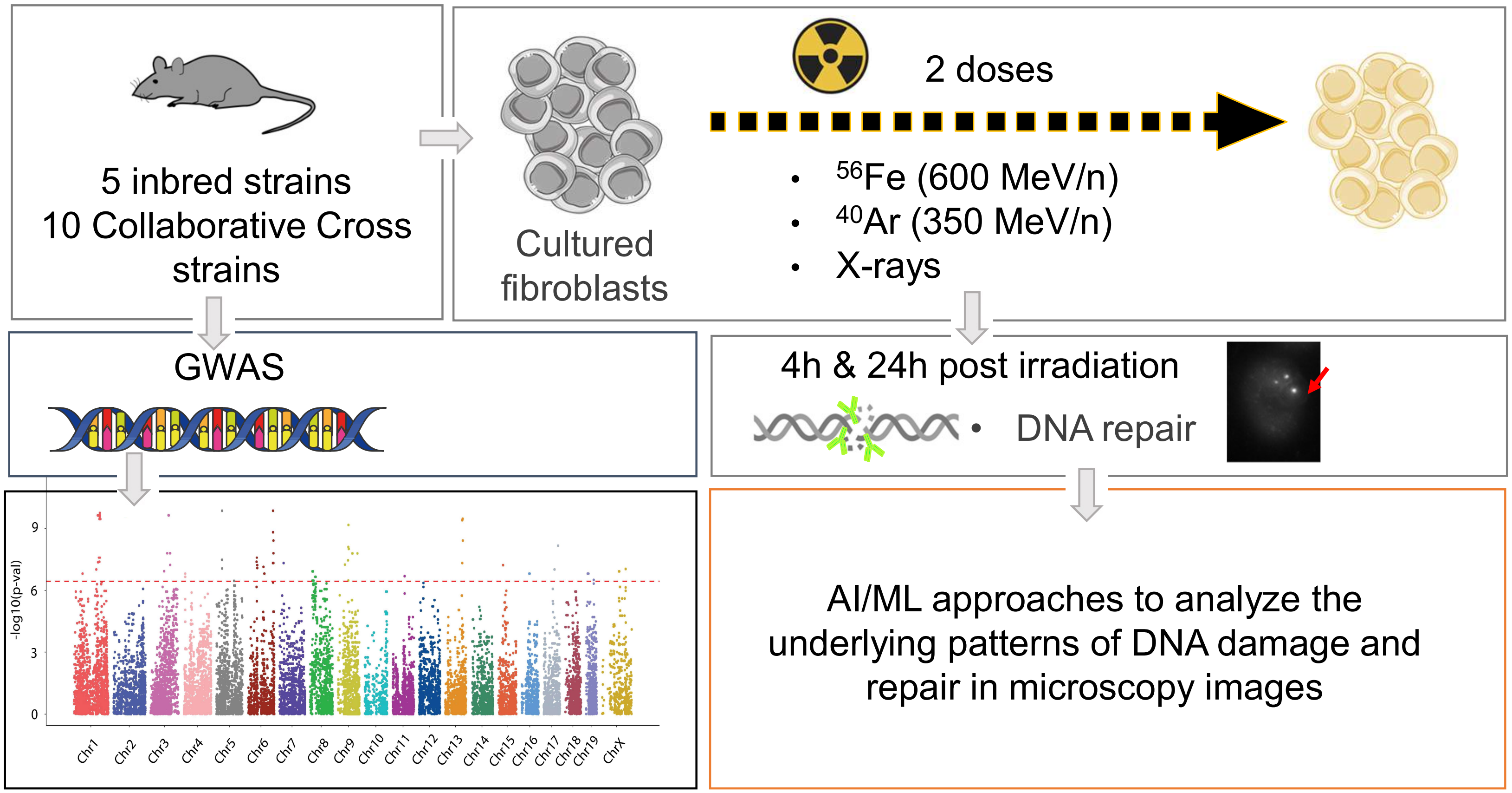




Example: Barcellos-Hoff Lab low dose data

Radiation data available in OSDR: rodent

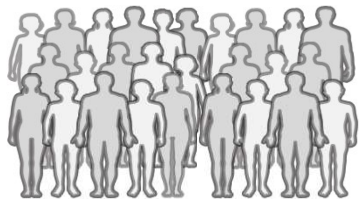
OSD-366



Cekanaviciute et al., Life Sci Space Res 2022

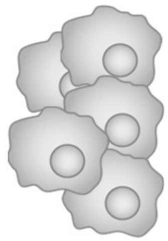
Radiation data available in OSDR: human

OSD-408

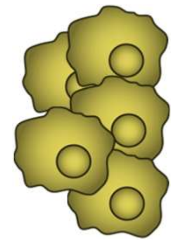


761 healthy donors
18-75 years old
1:1 male:female

Whole genome sequencing



Blood immune cells



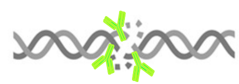
- ^{56}Fe, 600 MeV/n
- ^{40}Ar, 350 MeV/n
- ^{28}Si, 350 MeV/n
- Gamma rays

Particles with increasing energy

Control

Outcomes: baseline, 4h & 24h post irradiation

- DNA repair foci



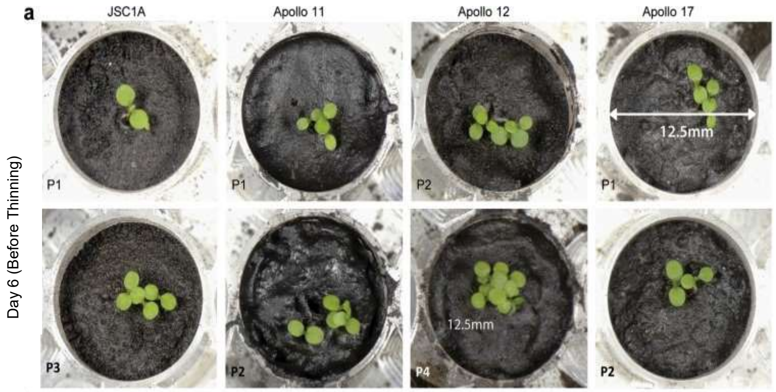
AI/ML approaches to analyze the underlying patterns of DNA damage and repair in microscopy images

Plants grown in Apollo lunar regolith present stress-associated transcriptomes that inform prospects for lunar exploration.

OSD-476

Arabidopsis thaliana germinates and grows in regolith

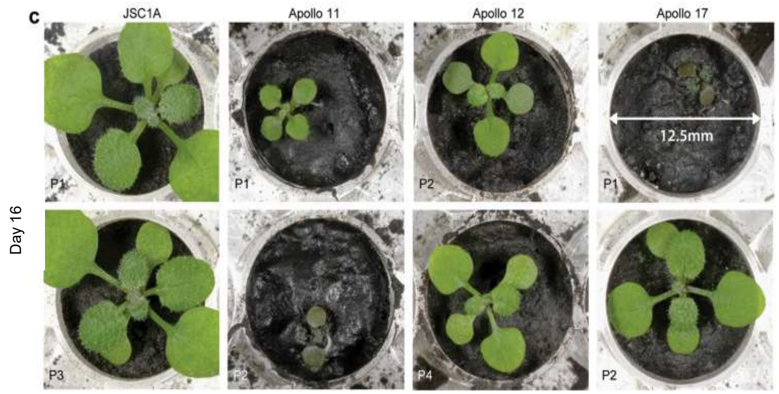
- Regolith samples from Apollo 11, 12, and 17 were used.
- JSC-1A = simulant control.
- Germination rates were close to 100% in all sources of Apollo regolith.
- Regolith-grown seedlings, however, did not thrive compared to JSC-1A controls.



Day 6

Plant transcriptomes differ by morphology

- Plants with a “severe” phenotype (tiny, abnormal morphology and reddish-black pigmentation) had 1000+ DEGs, demonstrating severe reactions to the regolith.
- All regolith samples, irrespective of Apollo site, significantly evoked differentially expressed genes (DEGs) indicative of a strong stress response.
- 71% of the DEGs typically were associated with salt, metal, and reactive oxygen species stress.

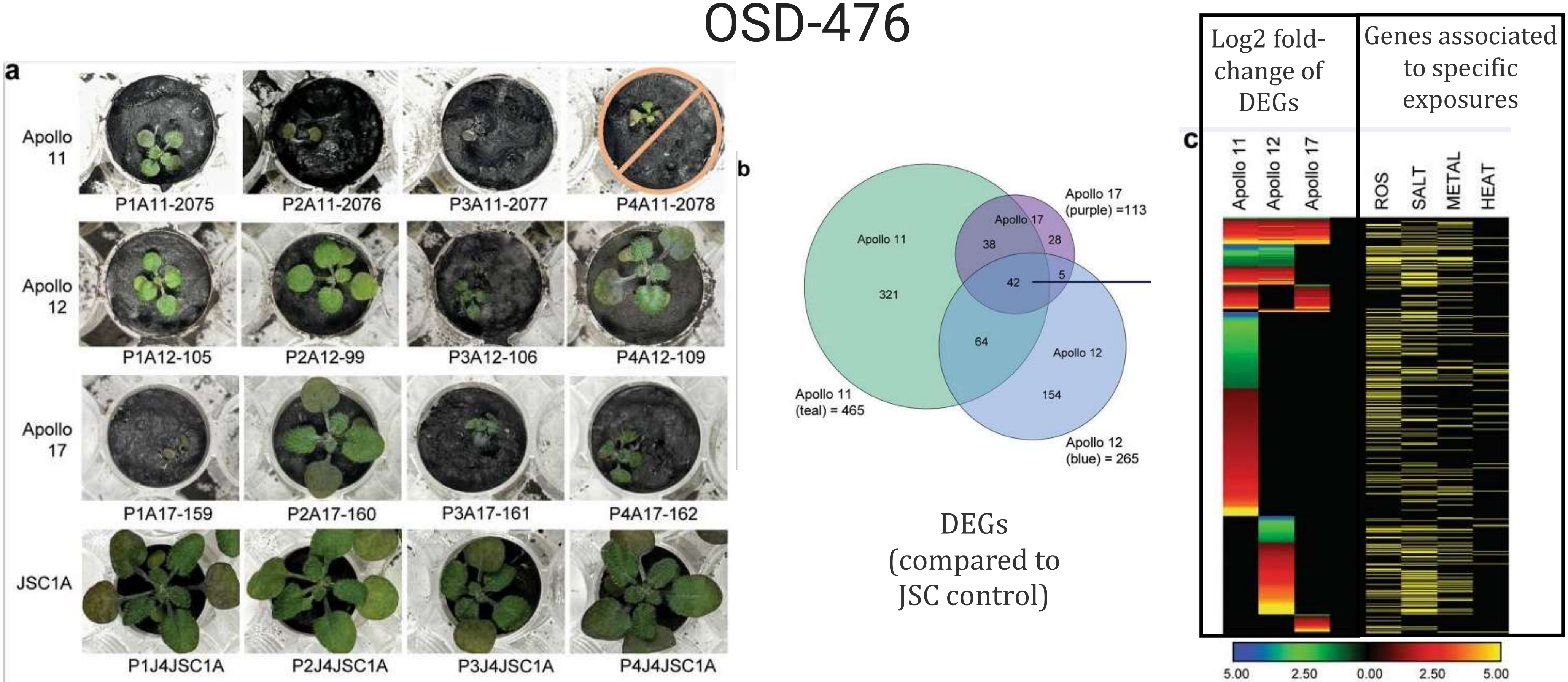


Day 16

Paul AL, et al 2022 Commun Biol. 2022 May 12;5(1):382.

Plants grown in Apollo lunar regolith present stress-associated transcriptomes that inform prospects for lunar exploration.

OSD-476



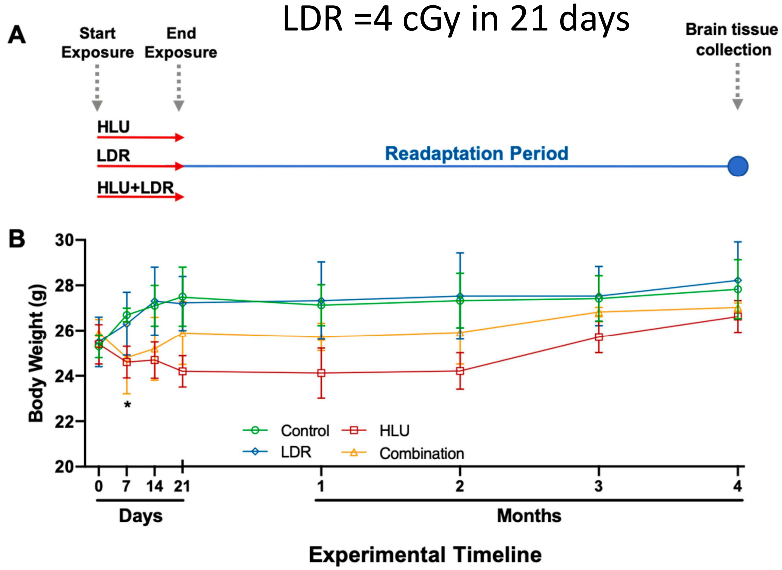
Paul AL, et al 2022 Commun Biol. 2022 May 12;5(1):382.

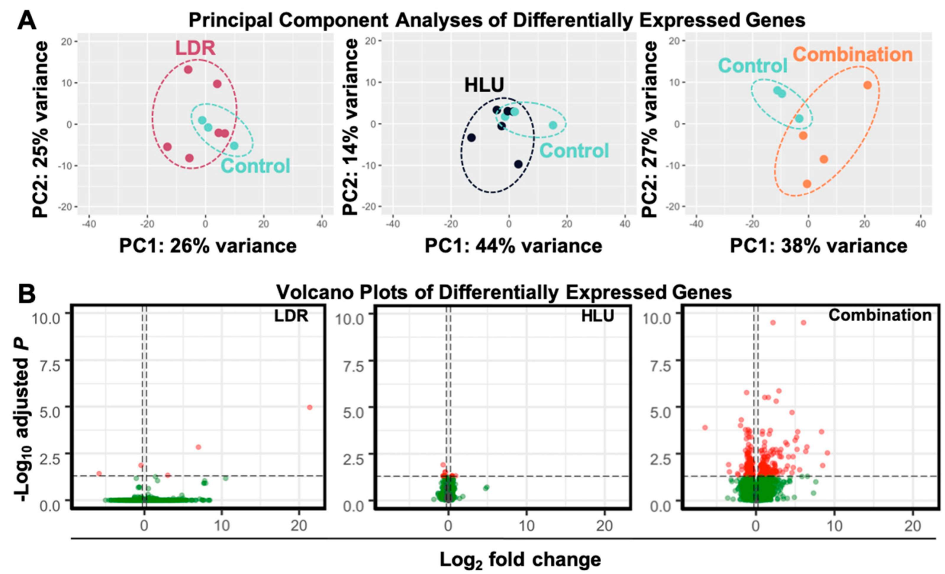
Combined effects of radiation and unloading OSD-202

Mice Exposed to Combined Chronic Low-Dose Irradiation and Modeled Microgravity Develop Long-Term Neurological Sequelae

by Eliah G. Overbey [1,†], Amber M. Paul [2,3,†], Willian A. da Silveira [4], Candice G.T. Tahimic [2,5], Sigrid S. Reinsch [2], Nathaniel Szewczyk [6], Seta Stanbouly [7], Charles Wang [8,9], Jonathan M. Galazka [2,*] and Xiao Wen Mao [7,*]

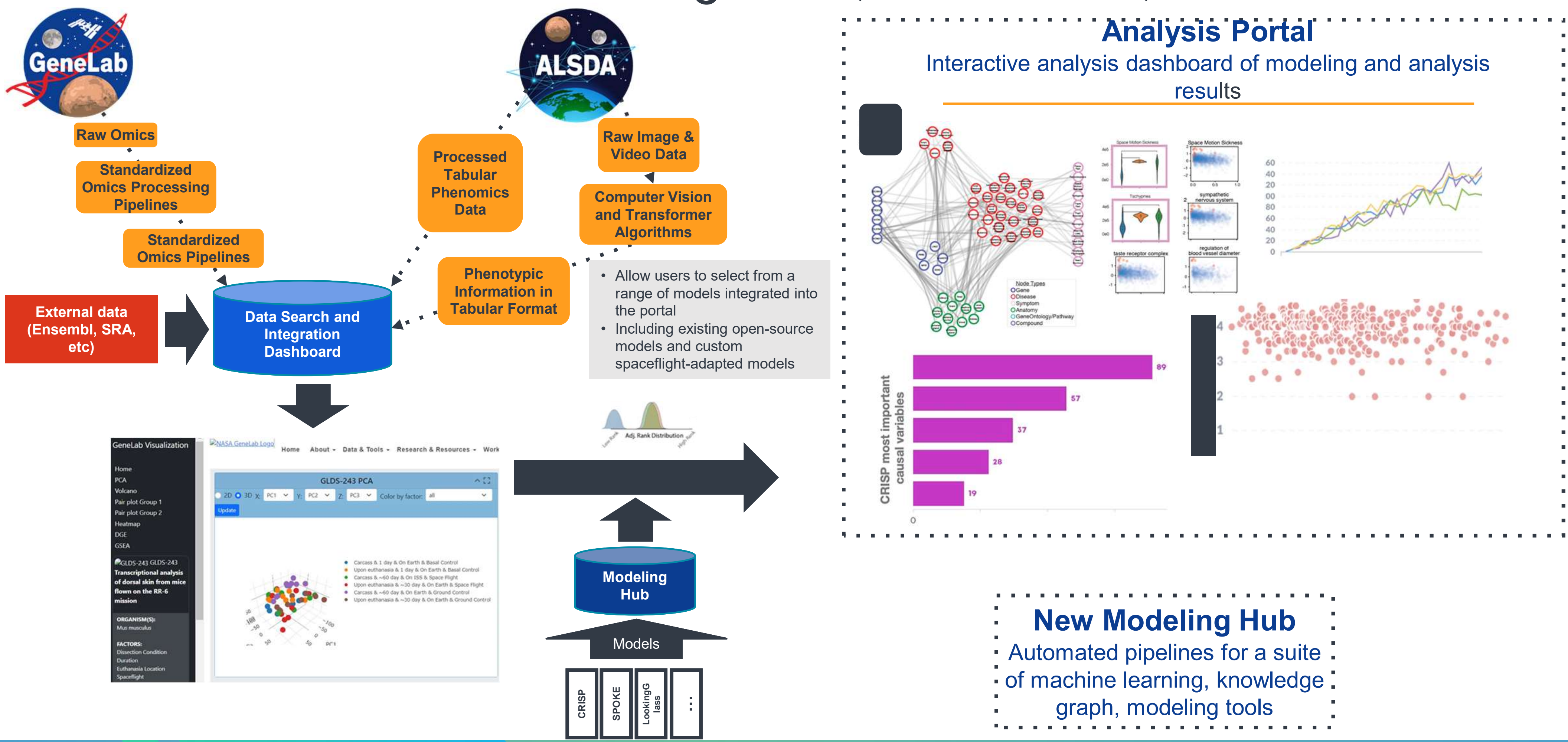
LDR =4 cGy in 21 days



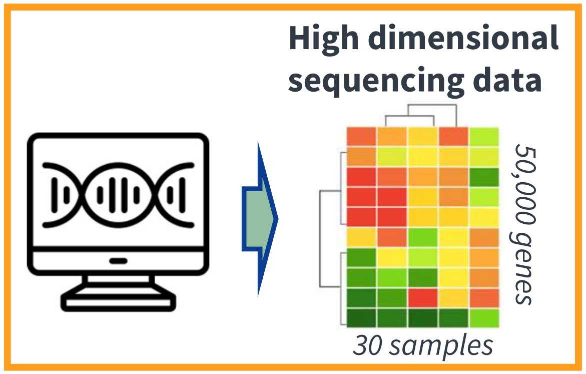


Int. J. Mol. Sci. **2019**, *20*(17), 4094; https://doi.org/10.3390/ijms20174094

Future: Streamlined Data Integration, Multimodal, Multihierarchical



Space biological data analysis challenges



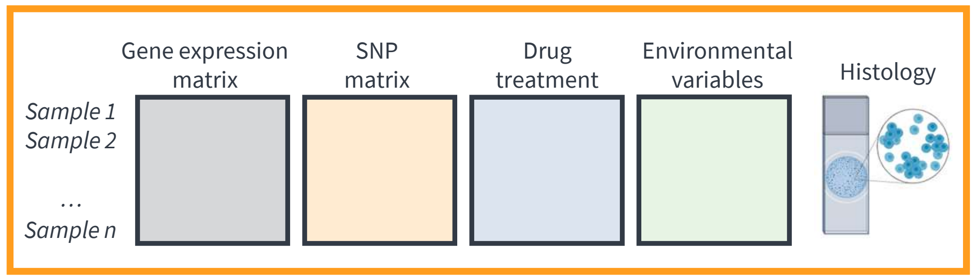
Space Biological Data Challenges

- **Small sample *n***
- **High feature count**
- Heterogeneous data
- Sparse data
- Transfer from model to human

Space biological data analysis challenges

Space Biological Data Challenges

- Small sample *n*
- High feature count
- **Heterogeneous data**
- **Sparse data**
- Transfer from model to human



Space biological data analysis challenges

Space Biological Data Challenges

- Small sample *n*
- High feature count
- Heterogeneous data
- Sparse data
- **Transfer from model to human**

Complex (Biological) Systems need Complex Models

Multiple approaches for characterizing patterns in biology

STATISTICAL METHODS

- Draw conclusions from observed data (**inference**)
- Assume specific data distributions
- Examples: hypothesis testing, correlative analysis

MACHINE LEARNING

- Learn from data to make predictions on unseen data (**prediction**)
- Able to model nonlinear relationships without assuming a data distribution
- Examples: classification, regression, clustering

Bzdok et al., *Nature Methods* 2018

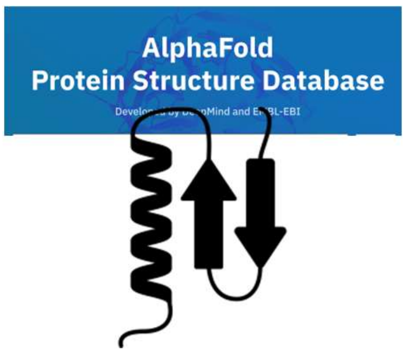
ML Learns and Predicts Complex Biological Phenomena

J Thorac Imaging. 2020 Nov 1;35(6):361-368. doi: 10.1097/RTI.0000000000000544.

A Novel Machine Learning-derived Radiomic Signature of the Whole Lung Differentiates Stable From Progressive COVID-19 Infection: A Retrospective Cohort Study

Liping Fu [1], Yongchou Li [2], Aip

The opportunity: adapt ML principles to power knowledge discovery and address key challenges in space biological research



Sci Data. 2021 Apr 29;8(1):121. doi: 10.1038/s41597-021-00900-3.

COVID-CT-MD, COVID-19 computed tomography scan dataset applicable in machine learning and deep learning

Parnian Afshar [1], Shahin Heidarian [2], Nastaran Enshaei [1], Farnoosh Naderkhani [1], Moezedin Javad Rafiee [3], Anastasia Konstantinos N Plataniotis [7], Arash

PLoS One. 2013 Apr 30;8(4):e61318. doi: 10.1371/journal.pone.0061318. Print 2013.

Machine learning prediction of cancer cell sensitivity to drugs based on genomic and chemical properties

Michael P Menden [1], Francesco Iorio, Mathew Garnett, Ultan McDermott, Cyril H Benes, Pedro J Ballester, Julio Sae

6th International Conference on Smart Computing and Communications, ICSCC 2017, 7-8 December 2017, Kurukshetra, India

Lung Cancer Detection using CT Scan Images

Suren Makaju[a], P.W.C. Prasad

Review

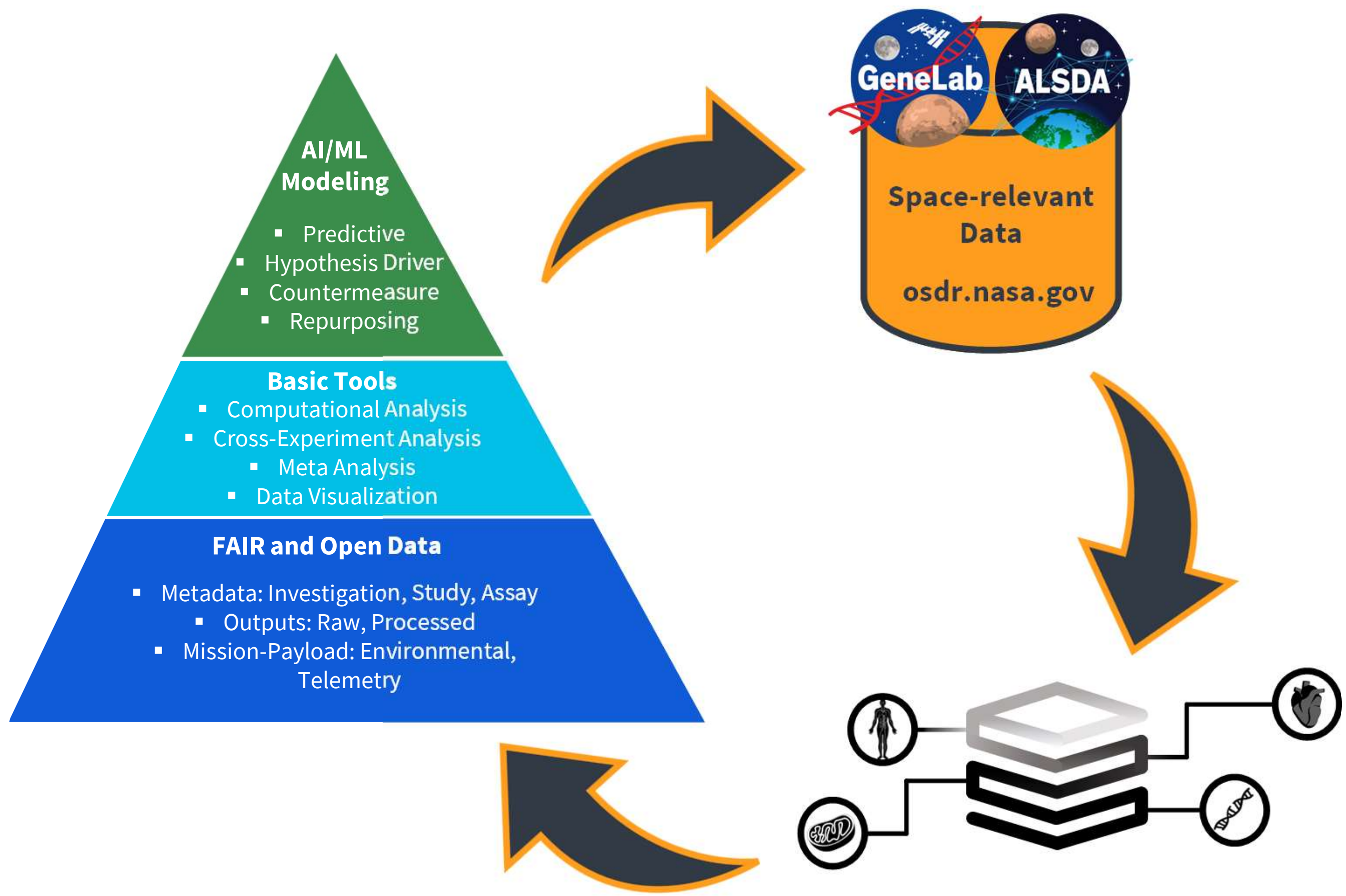
Iran J Public Health. 2017 Feb;46(2):165-172.

Improving the Prediction of Survival in Cancer Patients by Using Machine Learning Techniques: Experience of Gene Expression Data: A Narrative Review

Azadeh Bashiri [1], Marjan Ghazisaeedi [1], Reza Safdari [1], Leila Shahmoradi [1], Hamide Ehtesham [1]

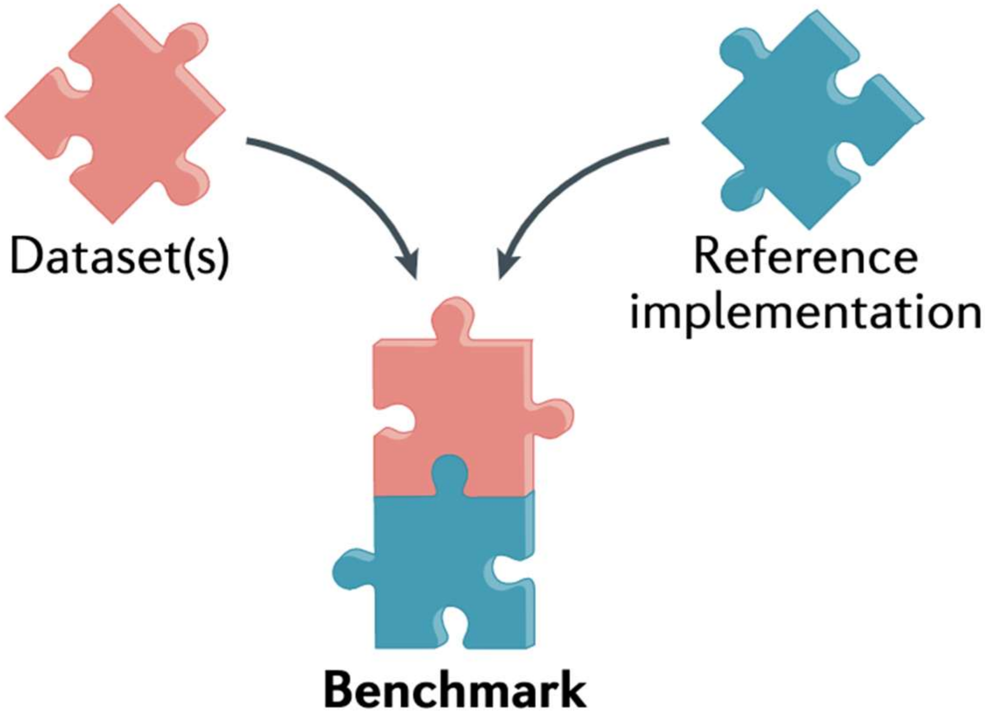
AI for Life in Space working group: AI4LS

Leveraging ML and AI methods to model space biology data from the NASA Open Science Data Repository: **NASA GeneLab** (omics) and **NASA Ames Life Sciences Data Archive** (ALSDA; phen-omics) to better understand the complex effects of spaceflight on living systems across hierarchical biological levels.



Benchmark Datasets for Space Biology

- **Scientific ML benchmarking**—Best ML algorithm for this problem
- **Application benchmarking**—Algorithm performance
- **System benchmarking**—Hardware and software architecture

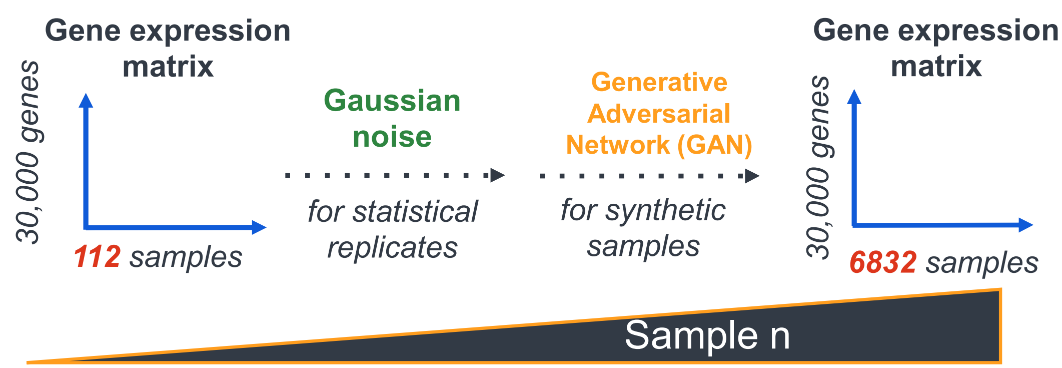


Thiyagalingam et al, *Nature Reviews* 2022

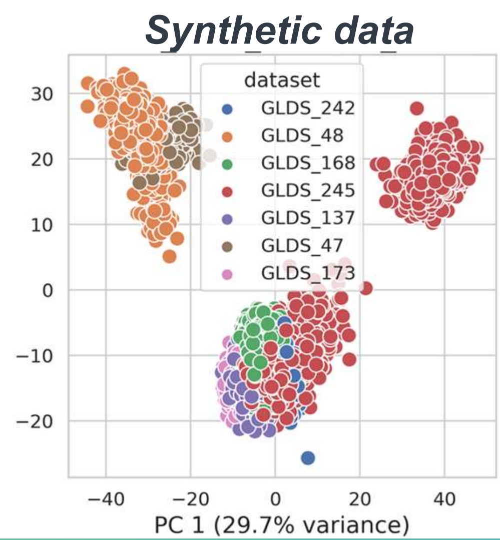
RNA sequencing benchmark dataset

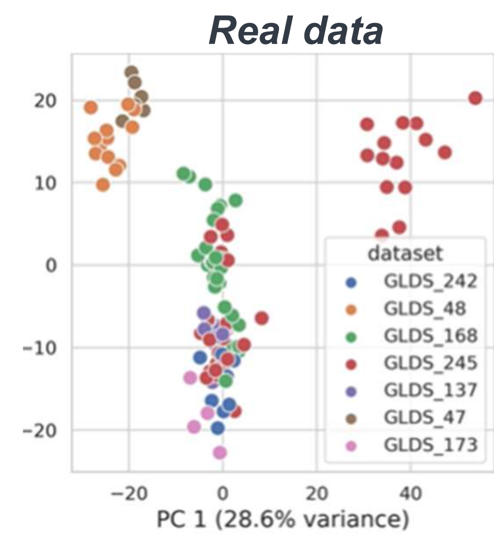
Scientific motivation: Effects of spaceflight on mouse liver health

AI readiness pipeline:



GeneLab Data Set	Tissue	Spaceflight Mission
47	liver	RR1 CASIS
48	liver	RR1 NASA
168	liver	RR1 NASA RR3 CASIS
137	liver	RR3 CASIS
173	liver	STS-135
242	liver	RR9
245	liver	RR6



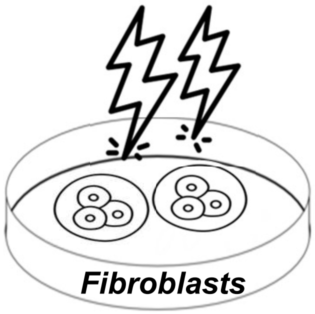


Fluorescence microscopy benchmark dataset

Scientific motivation: simulated space radiation causes cellular DNA damage

15 mouse strains

Simulated galactic cosmic rays

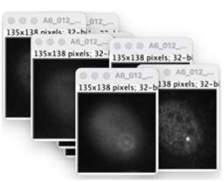


- 53PB1+ immunocytochemistry
- High-throughput imaging
- Automated quantification of 53 PB1+ radiation-induced foci

Penninckx et al. *Radiation Research* 2019

aws

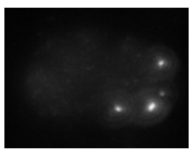
AI readiness pipeline:



maximum intensity projection

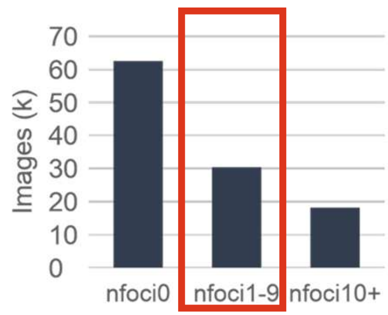
16-bit conversion

Raw Dataset (n = 94,193):
32-bit Z stacks (9 indices)

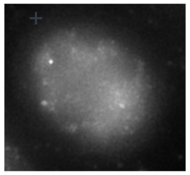


Max Intensity Dataset (n=94,193):
16-bit single-index TIFFs

automatically estimated nfoci



nfoci10++



Labels:

- *nfoci* (number of 53BP1+ DNA damage foci)
- *Radiation* (X-ray, ^{56}Fe…)

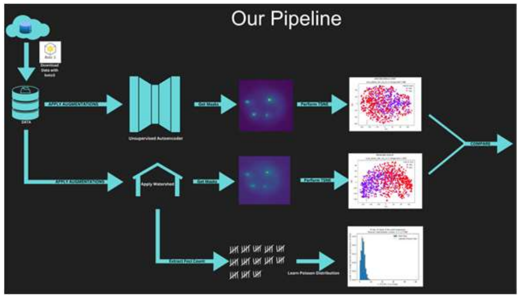
Cloud-based ML-ready Data Increase Scientific Community Engagement

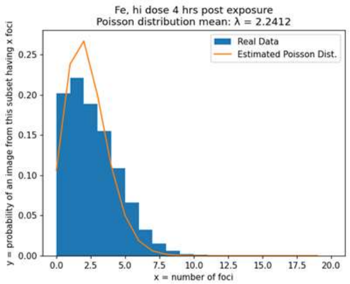
UC Irvine CS175: "Project in Artificial Intelligence" Senior Course

- BPS Microscopy benchmark dataset formed the basis for UCI CS175 senior ML projects
- Real-world data and scientific problems inspired the students to generate creative solutions
- ML-ready dataset allowed students to spend time on ML rather than preprocessing

- 9 teams focused on a variety of scientific questions:
  - Supervised classification
  - Unsupervised learning
  - Self-supervised learning
  - Segmentation and detection
  - Graph neural networks
  - Generating synthetic data

Image Segmentation and Foci Detection





Large Language Model for NASA Science

- NASA Science Mission Directorate and IBM working to train a large language model (LLM) for NASA science
- Training on NASA scientific text data from the 5 scientific divisions
  - Biological and Physical Sciences providing relevant full-text papers and abstracts
- The encoder model will be made public, and the generative model will be internal NASA use only
- LLM performance will be validated via multiple mechanisms including a question and answer (Q&A) task similar to the [Stanford SQuAD dataset]
- Soliciting Q&A tasks from subject matter experts in space relevant biology
  - Question
  - Correct answer
  - Source paragraph/sentence from literature/website for correct answer
  - URL to source
- Form to submit Q&A tasks: https://forms.gle/qtX2wSCnLEBqzMRFA

Form for Q&A Tasks

Large Language Model for NASA Science

Example Q&A Task

Question: Define a DNA double strand break

Correct answer: DNA DSBs consist of two SSBs (single DNA strand breaks) within one helical turn on opposing strands of DNA.

URL to source document: https://pubmed.ncbi.nlm.nih.gov/23311752/

Paragraph from source document where correct answer would be found: DNA DSBs consist of two SSBs within one helical turn on opposing strands of DNA. IR induced DNA DSBs are not usually simple lesions but rather contain multiple complex lesions and overhanging ends (or ragged ends) of DNA that cannot be ligated directly and therefore must be removed before repair can begin.

As you are designing your Q&A tasks, keep in mind that the models will be trained on the following data sources relevant to BPS science:

1. Full text articles from Radiation Research Society 2021-2023
2. All abstracts from PubMed
3. Full text articles from PMC, subset with keywords: "space biology", "microgravity", "astronaut", "space radiation", "space microbiology", "space plant biology", "space animal biology", "space medicine", "radiation biology", "mechanical unloading", "partial weight-bearing", "gravitropism", "mechanosensing", "hypergravity", "life support"
4. GeneLab and Open Science Data Repository websites
5. BPS Taskbook
6. NASA Technical Reports Server
7. NASA Science Discovery Engine

Form for Q&A Tasks

Acknowledgments

Radiation Biophysics Lab at NASA Ames Research Center:

Egle Cekanaviciute, Estrella Passerat de la Chapelle, Ananya Bansal

University of Nevada, Reno – Tin Nguyen Lab

AI4LS: Lauren Sanders, James Casaletto

RadLab Team: Kirill Grigorev, Jack Miller, Livio Narici, Lauren Sanders, Ana Uriarte Acuna

The entire OSDR team

Mason Lab at Cornell University

Chris Mason, Cem Meydan, Jonathan Foox

Brookhaven National Laboratory

Adam Rusek, Peter Guida

Space Biosciences
NASA AMES RESEARCH CENTER

NASA FUNDING:
NNJ16HP24I

FORMER COLLABORATORS
LBNL: Gary Karpen, Jian-Hua Mao, Antoine Snijders
NASA: Steve Blattnig, Artem Ponomarev, Ianik Plante
CSU: Michael Weil
UCSF: Mary Helen Barcellos-Hoff
INSERM, France: François Paris
Jules Bordet Hospital, Belgium: Sébastien Penninckx
Exogen Biotechnology Inc.: Jonathan Tang, Antony S. Tin